

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2005, 19:17:09 ; Search time 48056 Seconds  
(without alignments)  
7167.048 Million cell updates/sec

Title: US-10-673-885-1  
Perfect score: 7108  
Sequence: 1 taattgtacttgcagaa.....gcttctctactgtgcttt 7108

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hlg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_ey.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID          | Description        |
|------------|--------|-------------|--------|-------------|--------------------|
| 1          | 7108   | 100.0       | 7108   | 6 AR527423  | AR527423 Sequence  |
| 2          | 6800.8 | 95.7        | 6897   | 6 AX751409  | AX751409 Sequence  |
| 3          | 6799.8 | 95.7        | 7225   | 6 AX644236  | AX644236 Sequence  |
| 4          | 6683.6 | 94.0        | 6903   | 6 AX686453  | AX686453 Sequence  |
| 5          | 6453   | 90.8        | 6994   | 6 AX686449  | AX686449 Sequence  |
| 6          | 5619.4 | 79.1        | 5877   | 6 AX751415  | AX751415 Sequence  |
| 7          | 5294.2 | 74.5        | 7396   | 10 AF063249 | AF063249 Rattus no |
| 8          | 2618.6 | 36.8        | 3973   | 6 AR073855  | AR073855 Sequence  |
| 9          | 2618.6 | 36.8        | 3973   | 6 I32039    | I32039 Sequence 21 |
| 10         | 2551   | 35.9        | 2565   | 6 AX686451  | AX686451 Sequence  |
| 11         | 2251   | 31.7        | 3149   | 6 CQ726350  | CQ726350 Sequence  |
| 12         | 1118.8 | 15.7        | 2692   | 6 AR073854  | AR073854 Sequence  |
| 13         | 1118.8 | 15.7        | 2692   | 6 I32038    | I32038 Sequence 14 |
| 14         | 1071.6 | 15.1        | 2309   | 6 AR073853  | AR073853 Sequence  |
| 15         | 1071.6 | 15.1        | 2309   | 6 I32037    | I32037 Sequence 10 |
| 16         | 905.2  | 12.7        | 990    | 6 AX751414  | AX751414 Sequence  |
| 17         | 789.6  | 11.1        | 2173   | 6 AR073851  | AR073851 Sequence  |
| 18         | 789.6  | 11.1        | 2173   | 6 I32035    | I32035 Sequence 6  |
| 19         | 675.8  | 9.5         | 834    | 6 CQ731368  | CQ731368 Sequence  |

|    |       |     |        |             |                      |
|----|-------|-----|--------|-------------|----------------------|
| 20 | 532   | 7.5 | 532    | 6 AR073852  | AR073852 Sequence    |
| 21 | 532   | 7.5 | 532    | 6 I32036    | I32036 Sequence 8    |
| 22 | 401.6 | 5.6 | 402    | 9 AF169351  | AF169351 Homo sapi   |
| 23 | 330.4 | 4.6 | 163912 | 9 AC074031  | AC074031 Homo sapi   |
| 24 | 329.2 | 4.6 | 659    | 6 CQ723010  | CQ723010 Sequence    |
| 25 | 319.4 | 4.5 | 321    | 6 AR073850  | AR073850 Sequence    |
| 26 | 319.4 | 4.5 | 321    | 6 I32034    | I32034 Sequence 3    |
| 27 | 315   | 4.4 | 254366 | 6 AR527424  | AR527424 Sequence    |
| 28 | 285.6 | 4.0 | 179162 | 2 AC132076  | AC132076 Bos tauru   |
| 29 | 285.6 | 4.0 | 236486 | 2 AC138166  | AC138166 Bos tauru   |
| 30 | 283   | 4.0 | 162738 | 9 AC083812  | AC083812 Homo sapi   |
| 31 | 277   | 3.9 | 174126 | 2 AC135552  | AC135552 Canis fam   |
| 32 | 276   | 3.9 | 141299 | 2 AC132080  | AC132080 Felis cat   |
| 33 | 267.4 | 3.8 | 154909 | 2 AC136206  | AC136206 Sus scrof   |
| 34 | 266.8 | 3.8 | 363    | 6 CQ736545  | CQ736545 Sequence    |
| 35 | 251.2 | 3.5 | 162738 | 9 AC083812  | AC083812 Homo sapi   |
| 36 | 247.2 | 3.5 | 230799 | 2 AC098397  | AC098397 Rattus no   |
| 37 | 245.6 | 3.5 | 203839 | 10 AC021642 | AC021642 Mus muscu   |
| 38 | 230.8 | 3.2 | 412    | 6 AX751416  | AX751416 Sequence    |
| 39 | 225.6 | 3.2 | 164493 | 2 AC078825  | AC078825 Homo sapi   |
| 40 | 214.6 | 3.0 | 4053   | 5 GGU65891  | GGU65891 Gallus gall |
| 41 | 214   | 3.0 | 214819 | 10 AC123948 | AC123948 Mus muscu   |
| 42 | 201.6 | 2.8 | 261708 | 2 AC098350  | AC098350 Rattus no   |
| 43 | 195.2 | 2.7 | 3581   | 4 OC032587  | OC032587 Oryctolagus |
| 44 | 195.2 | 2.7 | 5704   | 4 OC009490  | OC009490 Oryctolagus |
| 45 | 194.6 | 2.7 | 5515   | 5 GGA238216 | GGA238216 Gallus ga  |

ALIGNMENTS

RESULT 1  
AR527423  
LOCUS AR527423 7108 bp DNA linear PAT 08-OCT-2004  
DEFINITION Sequence 1 from patent US 6723547.  
ACCESSION AR527423  
VERSION AR527423.1 GI:53914473  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 7108)  
Beasley,E.M., Webster,M., Francesco,V.D. and Wei,M.-H.  
TITLE Isolated human phosphatase proteins, nucleic acid molecules  
encoding human phosphatase proteins, and uses thereof  
JOURNAL Patent: US 6723547-A 1 20-APR-2004;  
FEATURES  
source Location/Qualifiers  
1..7108  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

|                       |                 |   |           |              |
|-----------------------|-----------------|---|-----------|--------------|
| Query Match           | 100.0%;         | Score 7108;   | DB 6;     | Length 7108; |
| Best Local Similarity | 100.0%;         | Pred. No. 0;  |           |              |
| Matches 7108;         | Conservative 0; | Mismatches 0;   | Indels 0; | Gaps 0;      |
| QY                    | 1               | TAATTGTGACTTGCAGAGGATCTGCTTTAAATCATTTAATGCGAGGCAACATTCTCT     |           | 60           |
| Db                    | 1               | TAATTGTGACTTGCAGAGGATCTGCTTTAAATCATTTAATGCGAGGCAACATTCTCT     |           | 60           |
| QY                    | 61              | CTAGAGCCATCAATGTGATTTCTACTGCGTGAATAAATGTATAAAGATGGATTTCTTATC  |           | 120          |
| Db                    | 61              | CTAGAGCCATCAATGTGATTTCTACTGCGTGAATAAATGTATAAAGATGGATTTCTTATC  |           | 120          |
| QY                    | 121             | ATTTTCTTTTACTTTTATTTGGACTTCAGAGACACAGGTTGATGTTTCCATGCGTT      |           | 180          |
| Db                    | 121             | ATTTTCTTTTACTTTTATTTGGACTTCAGAGACACAGGTTGATGTTTCCATGCGTT      |           | 180          |
| QY                    | 181             | CCTGGTACTAGGTACGATATAACCATCTCTCAATTTCTACACATACACCTCACCCTGTT   |           | 240          |
| Db                    | 181             | CCTGGTACTAGGTACGATATAACCATCTCTCAATTTCTACACATACACCTCACCCTGTT   |           | 240          |
| QY                    | 241             | ACTAGAATAGTGACACCAAAATGTAAACAAACGAGGCGCTCCAGTCTTCTTCAGCCGGGAA |           | 300          |



QY 2461 CCCAATGGAATCATATAAAATAATATACAAATTTATCTCAAGAGAGATATGGAATGAGGAA 2520  
DB 2461 CCCAATGGAATCATATAAAATAATATACAAATTTATCTCAAGAGAGATATGGAATGAGGAA 2520  
QY 2521 AGAAGTATAATACAACTCTTTAAACCCAAACAAATTAAGTACTGGAAGAAATATACCCAA 2580  
DB 2521 AGAAGTATAATACAACTCTTTAAACCCAAACAAATTAAGTACTGGAAGAAATATACCCAA 2580  
QY 2581 TATATCAATGAGGTGCTGCTAGTACACTGAAAGGTGAAGAGTTCGGAGTCTCCCATTA 2640  
DB 2581 TATATCAATGAGGTGCTGCTAGTACACTGAAAGGTGAAGAGTTCGGAGTCTCCCATTA 2640  
QY 2641 AGTATACAGGAGGAGAGTCTCTGATTTCTCCCTCAAGACTTCTCTGTAAGAACAG 2700  
DB 2641 AGTATACAGGAGGAGAGTCTCTGATTTCTCCCTCAAGACTTCTCTGTAAGAACAG 2700  
QY 2701 TTGTCGTGTGTCAGGTGAGTGTGTCATGCGAACCCCTGGAGCCAAATGGAATTTATC 2760  
DB 2701 TTGTCGTGTGTCAGGTGAGTGTGTCATGCGAACCCCTGGAGCCAAATGGAATTTATC 2760  
QY 2761 CTTTATTTACACAGTTTATGTCGGAATAGATCATATTAATAAACTATTAATGTCACCTGAA 2820  
DB 2761 CTTTATTTACACAGTTTATGTCGGAATAGATCATATTAATAAACTATTAATGTCACCTGAA 2820  
QY 2821 ACATCATTTGAGTTATCAGATTTGGAATTAATGTTGAATACAGTGTCTTATGTAACAGCT 2880  
DB 2821 ACATCATTTGAGTTATCAGATTTGGAATTAATGTTGAATACAGTGTCTTATGTAACAGCT 2880  
QY 2881 AGCACAGATTTGTCGATGAGGAAACAGGAGCAATATCATTTCAACACCCAGAG 2940  
DB 2881 AGCACAGATTTGTCGATGAGGAAACAGGAGCAATATCATTTCAACACCCAGAG 2940  
QY 2941 GGAGCACCAAGCGATCTCCCAAGAGTGTATTAATGCAAACTCAGTCTCTCATCAATA 3000  
DB 2941 GGAGCACCAAGCGATCTCCCAAGAGTGTATTAATGCAAACTCAGTCTCTCATCAATA 3000  
QY 3001 ATTCTTTTCTGGACACCTCTTCAAACTTAATGGAATTAACAATTTACTCTGTTTAT 3060  
DB 3001 ATTCTTTTCTGGACACCTCTTCAAACTTAATGGAATTAACAATTTACTCTGTTTAT 3060  
QY 3061 TACAGAAATCTTCAGTACTTTTATGCAAGATTTTACACTCCATGACATTAACCAATGAC 3120  
DB 3061 TACAGAAATCTTCAGTACTTTTATGCAAGATTTTACACTCCATGACATTAACCAATGAC 3120  
QY 3121 TTTGACAAATGACTGTATCCAAATTAATAGATAAATCTGCAATATTTACACTACTATACA 3180  
DB 3121 TTTGACAAATGACTGTATCCAAATTAATAGATAAATCTGCAATATTTACACTACTATACA 3180  
QY 3181 TTTTGGTTAAACAGCAAGTACTTCAGTTGGAAATGGGAATAAAGCAGTGACATTTGAA 3240  
DB 3181 TTTTGGTTAAACAGCAAGTACTTCAGTTGGAAATGGGAATAAAGCAGTGACATTTGAA 3240  
QY 3241 GTATACACAGATCAAGACATACCTGAAGGTGTTGGAAACCTGACTTACGATTCATTT 3300  
DB 3241 GTATACACAGATCAAGACATACCTGAAGGTGTTGGAAACCTGACTTACGATTCATTT 3300  
QY 3301 TCGTCAACTGCAATAAATGTAAGTGGTCCCAAGGTCTCAACCAAGGTCTAGTCTTC 3360  
DB 3301 TCGTCAACTGCAATAAATGTAAGTGGTCCCAAGGTCTCAACCAAGGTCTAGTCTTC 3360  
QY 3361 TACTATGTTTCACTGATCTTACAGCAGACTCTCCGCAATGAGACCCCTCTGTTTACA 3420  
DB 3361 TACTATGTTTCACTGATCTTACAGCAGACTCTCCGCAATGAGACCCCTCTGTTTACA 3420  
QY 3421 TATGAGAGAGCATATATTTTGAATATCTGGAAAAATACACTGATTAATTAATAAAT 3480  
DB 3421 TATGAGAGAGCATATATTTTGAATATCTGGAAAAATACACTGATTAATTAATAAAT 3480  
QY 3481 ACTCCATCAAGAAAGGATTTCTGATACCTTACTACCTGCGGAGTATACATCAAGACT 3540  
DB 3481 ACTCCATCAAGAAAGGATTTCTGATACCTTACTACCTGCGGAGTATACATCAAGACT 3540

QY 3541 GAAGAAGATGTCCAGAAACTTCCAAATTAATCAACACTTTTAAAAACCTTCTCTTACC 3600  
DB 3541 GAAGAAGATGTCCAGAAACTTCCAAATTAATCAACACTTTTAAAAACCTTCTCTTACC 3600  
QY 3601 TCAGTTCTCTTATCATGCGGATCCCGAGTAAAGCCAAATGCGTCAATAATAAGTTATGAT 3660  
DB 3601 TCAGTTCTCTTATCATGCGGATCCCGAGTAAAGCCAAATGCGTCAATAATAAGTTATGAT 3660  
QY 3661 TTAACCTTTACAGGACCAATGAAATTAATCTTTCAATTAATCTCTGATAATACATAATA 3720  
DB 3661 TTAACCTTTACAGGACCAATGAAATTAATCTTTCAATTAATCTCTGATAATACATAATA 3720  
QY 3721 TTGGAAGAGCTTTGACATTTTACATTAATAGCTTTTTCCTGCGGCAAGAACTAGAAAA 3780  
DB 3721 TTGGAAGAGCTTTGACATTTTACATTAATAGCTTTTTCCTGCGGCAAGAACTAGAAAA 3780  
QY 3781 GGAAGTGTCTTCCAGTATTTCTTTCTTTTACACAGATGAGTCAAGTCCGTTAGCACT 3840  
DB 3781 GGAAGTGTCTTCCAGTATTTCTTTCTTTTACACAGATGAGTCAAGTCCGTTAGCACT 3840  
QY 3841 CCAAAAAATTTGACTTTTAAATCAACTGTACTTCAAGACTTTGTATGGCTGAAATGGAGCCCA 3900  
DB 3841 CCAAAAAATTTGACTTTTAAATCAACTGTACTTCAAGACTTTGTATGGCTGAAATGGAGCCCA 3900  
QY 3901 AGTCTCTTCCAGGTGGTATTTGTAAGTATATAGTTTAAATTTCAATGCAACATGAACT 3960  
DB 3901 AGTCTCTTCCAGGTGGTATTTGTAAGTATATAGTTTAAATTTCAATGCAACATGAACT 3960  
QY 3961 GACACTATATATTAAGAAATATATCAGGATTTTAAACCTGAAAGCAAACTTTGTTGACTG 4020  
DB 3961 GACACTATATATTAAGAAATATATCAGGATTTTAAACCTGAAAGCAAACTTTGTTGACTG 4020  
QY 4021 GAAACAGTCAAGCACTTCTATCCGTGTATCTGCGTTTCAACAAAGTTGGAATGGCAAT 4080  
DB 4021 GAAACAGTCAAGCACTTCTATCCGTGTATCTGCGTTTCAACAAAGTTGGAATGGCAAT 4080  
QY 4081 CAATTTAGTAAATGTAAGAAATTTCAACCCAAAGATCAGTTCCAGATGTCGTGCAAGT 4140  
DB 4081 CAATTTAGTAAATGTAAGAAATTTCAACCCAAAGATCAGTTCCAGATGTCGTGCAAGT 4140  
QY 4141 ATGCAAGTCAAGCACTTCTGCGTGTATGTAAGTAAAGTGGATCCCAAGAAAG 4200  
DB 4141 ATGCAAGTCAAGCACTTCTGCGTGTATGTAAGTAAAGTGGATCCCAAGAAAG 4200  
QY 4201 GCAATGGAATTAATACGCAAGTATATGTAAGTAAAGTGGATCCCAAGAAAGTTCCT 4260  
DB 4201 GCAATGGAATTAATACGCAAGTATATGTAAGTAAAGTGGATCCCAAGAAAGTTCCT 4260  
QY 4261 CCCCAAGATCAGATGTACACTTTTCAAAAAGCTTTTCCCAATACCTCATATGCTTTTAAA 4320  
DB 4261 CCCCAAGATCAGATGTACACTTTTCAAAAAGCTTTTCCCAATACCTCATATGCTTTTAAA 4320  
QY 4321 GTAAGAGCTTCAACCTCAGCTGGTGAAGTGAAGCAAGTCCATGTCAGCACTA 4380  
DB 4321 GTAAGAGCTTCAACCTCAGCTGGTGAAGTGAAGCAAGTCCATGTCAGCACTA 4380  
QY 4381 CCTGAAACAGTTCCTCAGTGTTCCTCAAAATATTTGCTTTTCTGATGTTTCAAGTCACTAGT 4440  
DB 4381 CCTGAAACAGTTCCTCAGTGTTCCTCAAAATATTTGCTTTTCTGATGTTTCAAGTCACTAGT 4440  
QY 4441 GCAACATTCAGATGGAATAAGACCTGACATATCTTCTGGCTTACTTTTCAAAATTAACAAAT 4500  
DB 4441 GCAACATTCAGATGGAATAAGACCTGACATATCTTCTGGCTTACTTTTCAAAATTAACAAAT 4500  
QY 4501 ACCACTCACTTCTGCTCAAAATGCAAGTGGAAATCCGAAGATGTTGATAT 4560  
DB 4501 ACCACTCACTTCTGCTCAAAATGCAAGTGGAAATCCGAAGATGTTGATAT 4560  
QY 4561 CAAAAATTCATATACCTCTTATGAGTCACTTAACTGAAGAGACAGTATATGATTAAG 4620  
DB 4561 CAAAAATTCATATACCTCTTATGAGTCACTTAACTGAAGAGACAGTATATGATTAAG 4620  
QY 4621 AAATTTAGTGTATAGTTTCAAGTGGCTGCGAGCAACCAATGCTGGCTATGCAATGCT 4680







|||||  
Db 1441 TGGTCTGACCTTCACTTCACTTGCTACATTTATATATAACAGCCATCCAGATAAAACTTT 1500  
QY 1606 CTTGGAAGGAATAGAGCTGAAGACGACGCTTCAAGCTTGTAACTAACAAGGAATCAGTAT 1665  
Db 1501 CCTGCAAGGAATAGAGCTGAAGACGACGCTTCAAGCTTGTAACTAACAAGGAATCAGTAT 1560  
QY 1666 ATTACTGACATTCGACGCTGAACAGCTGCTTATGTATCAGGAGAGCTTGTACTTTCACT 1725  
Db 1561 ATTACTGACATTCGACGCTGAACAGCTGCTTATGTATCAGGAGAGCTTGTACTTTCACT 1620  
QY 1726 GAGCACATGATTAGTGTATCTGCTTTCAACCATCATGGGAGAAGGACCAACACAGTTCTC 1785  
Db 1621 GAGCACATGATTAGTGTATCTGCTTTCAACCATCATGGGAGAAGGACCAACACAGTTCTC 1680  
QY 1786 AGTGTTAGGACAGCTCAGCAGAGTGGCCAGCTCAATTAATAATTAATACTATAAAATATT 1845  
Db 1681 AGTGTTAGGACAGCTCAGCAGAGTGGCCAGCTCAATTAATAATTAATACTATAAAATATT 1740  
QY 1846 AGTTCCTTCATCTATTTTGTATATTCGGATCCTCCAGAAATATCCCAATGGAAAAATACT 1905  
Db 1741 AGTTCCTTCATCTATTTTGTATATTCGGATCCTCCAGAAATATCCCAATGGAAAAATACT 1800  
QY 1906 CACTATACGATTTATGCAATGGAAATGGATACAAACAGAGCAFTCCAGATAACTACCAT 1965  
Db 1801 CACTATACGATTTATGCAATGGAAATGGATACAAACAGAGCAFTCCAGATAACTACCAT 1860  
QY 1966 GATAACAGCTTTCTCATACAGGGTTAAAGAAATACAAATAACAAATAGAGTGGCA 2025  
Db 1861 GATAACAGCTTTCTCATACAGGGTTAAAGAAATACAAATAACAAATAGAGTGGCA 1920  
QY 2026 GCCTCAACCCAGATCGGAGAGTCTTCTGCTGGAAGAAATGACATCTTTGTGAACT 2085  
Db 1921 GCCTCAACCCAGATCGGAGAGTCTTCTGCTGGAAGAAATGACATCTTTGTGAACT 1980  
QY 2086 TCAGAGAGATGAACCGGAATCATCACTCAAGATGTGCAAGTAAATGATGTACCGCAGAT 2145  
Db 1981 TCAGAGAGATGAACCGGAATCATCACTCAAGATGTGCAAGTAAATGATGTACCGCAGAT 2040  
QY 2146 GAAATAGGTTGAAGTGGTCCACCCGAAAGCCCAATGGGATCAATATGCTTTATGAA 2205  
Db 2041 GAAATAGGTTGAAGTGGTCCACCCGAAAGCCCAATGGGATCAATATGCTTTATGAA 2100  
QY 2206 GTGCTATATAAATAATAGTACTTTATATATCAAGACACATCAACACAGACATAATA 2265  
Db 2101 GTGCTATATAAATAATAGTACTTTATATATCAAGACACATCAACACAGACATAATA 2160  
QY 2266 TTAAGGAACTTAAGACCTCACACCCCTATAPACATTTCTGTAAGGTCTTACACAGATTT 2325  
Db 2161 TTAAGGAACTTAAGACCTCACACCCCTATAPACATTTCTGTAAGGTCTTACACAGATTT 2220  
QY 2326 GGTCAATGGCAATCAGGTATCTTTTACTCTCTGTAAGGACTTCGAGAGCTGTGCTGTAT 2385  
Db 2221 GGTCAATGGCAATCAGGTATCTTTTACTCTCTGTAAGGACTTCGAGAGCTGTGCTGTAT 2280  
QY 2386 AGTGACCCGAAATATCACTTACAAAATATTTCTCTGGAGAGATCGAGCTATCATTC 2445  
Db 2281 AGTGACCCGAAATATCACTTACAAAATATTTCTCTGGAGAGATCGAGCTATCATTC 2340  
QY 2446 CTTTCCCCCAAGTAGTCCCAATGGAATGAAATATAAATAATATACATTTATCTCAAGAGAAGT 2505  
Db 2341 CTTTCCCCCAAGTAGTCCCAATGGAATGAAATATAAATAATATACATTTATCTCAAGAGAAGT 2400  
QY 2506 AATGGAAATAGGAAAGAACTATAATAACAACCTCTTTAAACCAAAACATTAAGTACTG 2565  
Db 2401 AATGGAAATAGGAAAGAACTATAATAACAACCTCTTTAAACCAAAACATTAAGTACTG 2460  
QY 2566 AAGAAATATACCAATATATCATTTAGGCTGCTGCTAGTACACTGAAAGGTGAAGAGTT 2625  
Db 2461 AAGAAATATACCAATATATCATTTAGGCTGCTGCTAGTACACTGAAAGGTGAAGAGTT 2520  
QY 2626 CGGAGTGTCCCATAAAGTATACCTGACGGAGGAAGATGCTCCTGATTTCTCCCTCCCAAGAC 2685

Db 2521 CGGAGTGTCCCATAAAGTATATCTGACGGAGGAAGATGCTCTGATTTCTCCCTCCCAAGAC 2580  
QY 2686 TTTCTCTGTAARACAGTTGCTGCTGTGTCACGGTGAAGTTGTCATGCGCAACACCCCTGGAG 2745  
Db 2581 TTTCTCTGTAARACAGTTGCTGCTGTGTCACGGTGAAGTTGTCATGCGCAACACCCCTGGAG 2640  
QY 2746 CCAATAGGAATTTATCTTTATTAACACAGTTTATGTCTGGAATAGATCATCATTAATAAAT 2805  
Db 2641 CCAATAGGAATTTATCTTTATTAACACAGTTTATGTCTGGAATAGATCATCATTAATAAAT 2700  
QY 2806 ATTAATGTCTCACTGAAACATCATTTGAGTTATCAGATTTGGATTAATAATGTAATACAT 2865  
Db 2701 ATTAATGTCTCACTGAAACATCATTTGAGTTATCAGATTTGGATTAATAATGTAATACAT 2760  
QY 2866 GCTTATGTAACAGCTAGCACACAGATTTTGGTGTGGAAGAAACAGGAAGCAATATCATTAGC 2925  
Db 2761 GCTTATGTAACAGCTAGCACACAGATTTTGGTGTGGAAGAAACAGGAAGCAATATCATTAGC 2820  
QY 2926 TTTCAAAACACAGAGGGAGCACCAAGCGATCTCTCCAAAGATGTTTATTTATGCAACCTC 2985  
Db 2821 TTTCAAAACACAGAGGGAGCACCAAGCGATCTCTCCAAAGATGTTTATTTATGCAACCTC 2880  
QY 2986 AGTTCCTTCATCAATTAATTTCTTTCTGGACACCTCTTCAAAACCTTAATGGGATTAACAA 3045  
Db 2881 AGTTCCTTCATCAATTAATTTCTTTCTGGACACCTCTTCAAAACCTTAATGGGATTAACAA 2940  
QY 3046 TATTACTCTGTTTATTAACAGAAATACTTCAAGTACTTTTATGCAAGATTTTATGCAACCTC 3105  
Db 2941 TATTACTCTGTTTATTAACAGAAATACTTCAAGTACTTTTATGCAAGATTTTATGCAACCTC 3000  
QY 3106 GAACTAAACCAATGACTTTGACAAATATGACTGTATCCAAATTTATAGATAAACTGCAATA 3165  
Db 3001 GAAGTAAACCAATGACTTTGACAAATATGACTGTATCCAAATTTATAGATAAACTGCAATA 3060  
QY 3166 TTTCACTACTATACATTTTGGTTTAAACAGAGTACTTTCAAGTGTGGAATGGGAATAAAGC 3225  
Db 3061 TTTCACTACTATACATTTTGGTTTAAACAGAGTACTTTCAAGTGTGGAATGGGAATAAAGC 3120  
QY 3226 AGTGACATCATTTGAGATATACACAGATCAAGACATACCTGGAAGGTTTCTGGAACCTG 3285  
Db 3121 AGTGACATCATTTGAGATATACACAGATCAAGACATACCTGGAAGGTTTCTGGAACCTG 3180  
QY 3286 ACTTACGAATCCATTTCTGTCATGCAATTAATGTAAGCTGGGTCCACCGGCTCAACCA 3345  
Db 3181 ACTTACGAATCCATTTCTGTCATGCAATTAATGTAAGCTGGGTCCACCGGCTCAACCA 3240  
QY 3346 AACGGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGAGACTCTCCGATGTGAGA 3405  
Db 3241 AACGGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGAGACTCTCCGATGTGAGA 3300  
QY 3406 CCACCTCTGTTACATATCAGAGAGCATATATTTTGAATACTGGAATAATACACTGAT 3465  
Db 3301 CCACCTCTGTTACATATCAGAGAGCATATATTTTGAATACTGGAATAATACACTGAT 3360  
QY 3466 TATATATTAATAATTTACTCCATCAACAGAAAGGGATTTCTGATACCTATATCTGCCAG 3525  
Db 3361 TATATATTAATAATTTACTCCATCAACAGAAAGGGATTTCTGATACCTATATCTGCCAG 3420  
QY 3526 CTATACATCAAGACTGGAAGAGTCCAGAAAATTTCAACCAATTAATCAACACTTTTAA 3585  
Db 3421 CTATACATCAAGACTGGAAGAGTCCAGAAAATTTCAACCAATTAATCAACACTTTTAA 3480  
QY 3586 AACCTTTCTCTACCTCAGTCTCTTATCATGGATCCCCAGTAAAGCAATGCTGCA 3645  
Db 3481 AACCTTTCTCTACCTCAGTCTCTTATCATGGATCCCCAGTAAAGCAATGCTGCA 3540  
QY 3646 ATAATAAGTTATGATTAATTTTCAAGGACCAAAATGAAAAATTTATTTCTTCACTTCT 3705  
Db 3541 ATAATAAGTTATGATTAATTTTCAAGGACCAAAATGAAAAATTTATTTCTTCACTTCT 3600  
QY 3706 GATAATTAATTAATGGAAGCTTTTCAACATTTATATATAGCTTTTGTGCTGCC 3765  
Db 3601 GATAATTAATTAATGGAAGCTTTTCAACATTTATATATAGCTTTTGTGCTGCC 3660

|    |      |   |      |
|----|------|---|------|
| Qy | 3766 | GCAAGAACTAGAAAAGGACTTGGTCCTTCCAGTATTCCTTTCTTTTACACAGATGAGTCA      | 3825 |
| Db | 3661 | GCAAGAACTAGAAAAGGACTTGGTCCTTCCAGTATTCCTTTCTTTTACACAGATGAGTCA      | 3720 |
| Qy | 3826 | GTGCCGTTAGCACTCCACAAAATTTGACATTTTAATCAACTGTACTTCAGACTTTGTATGG     | 3885 |
| Db | 3721 | GTGCCGTTAGCACTCCACAAAATTTGACATTTTAATCAACTGTACTTCAGACTTTGTATGG     | 3780 |
| Qy | 3886 | CTGAAATGGAGCCCAAGTCCTCTTCCAGGTGGTATGCTTAAAGATATATAGTTTTTAAATTT    | 3945 |
| Db | 3781 | CTGAAATGGAGCCCAAGTCCTCTTCCAGGTGGTATGCTTAAAGATATATAGTTTTTAAATTT    | 3840 |
| Qy | 3946 | CATGAACATGAATCTGACACTATATATATAAGATATATACAGATTTTAAACTGAGACC        | 4005 |
| Db | 3841 | CATGAACATGAACCTGACACTATATATATAAGAAATATATACAGATTTTAAACTGAGACC      | 3900 |
| Qy | 4006 | AAACTCTGTGGACTGGAAACCAAGTCAGCACCTACTCTATCCGTGTATCTCGTTTCACCCAAA   | 4065 |
| Db | 3901 | AAACTCTGTGGACTGGAAACCAAGTCAGCACCTACTCTATCCGTGTATCTCGTTTCACCCAAA   | 3960 |
| Qy | 4066 | GTGGAAATGGCAATCAATTTAGTAATGTAGTAAATTCACACCCAAAGAAATCAGTTTCCA      | 4125 |
| Db | 3961 | GTGGAAATGGCAATCAATTTAGTAATGTAGTAAATTCACACCCAAAGAAATCAGTTTCCA      | 4020 |
| Qy | 4126 | GATGTCGTGCAGATATGCGATGCAATGGCAACTAGCTGCGCAGTCAGTTTTAGTGAATGG      | 4185 |
| Db | 4021 | GATGTCGTGCAGATATGCGATGCAATGGCAACTAGCTGCGCAGTCAGTTTTAGTGAATGG      | 4080 |
| Qy | 4186 | GATCCACCCAAAAGGCAATTGAATAATAACGCAGTATATGTTAAAGCTTCAAGTTGAAAGGAAT  | 4245 |
| Db | 4081 | GATCCACCCAAAAGGCAATTGAATAATAACGCAGTATATGTTAAAGCTTCAAGTTGAAAGGAAT  | 4140 |
| Qy | 4246 | TCTACAAAAGTTTCTCCCAAGATCAATGTACACTTTTCAATAAGCTTCTTGGCCAAATACC     | 4305 |
| Db | 4141 | TCTACAAAAGTTTCTCCCAAGATCAATGTACACTTTTCAATAAGCTTCTTGGCCAAATACC     | 4200 |
| Qy | 4306 | TCATATGCTTTTAAAGTAAGAGCTTCAACCTCAGCTGCTCAAGTGAATGAAGCAATGC        | 4365 |
| Db | 4201 | TCATATGCTTTTAAAGTAAGAGCTTCAACCTCAGCTGCTCAAGTGAATGAAGCAATGC        | 4260 |
| Qy | 4366 | CATGTCAGCACACTACCTGAAAACAGTTTCCCAAGTGTTCGCCACAAATATGCTTTTCTGAT    | 4425 |
| Db | 4261 | CATGTCAGCACACTACCTGAAAACAGTTTCCCAAGTGTTCGCCACAAATATGCTTTTCTGAT    | 4320 |
| Qy | 4426 | GTTTCAGTCAACTAGTGCACCAATTCGACATGATGAATAGACTGACACTATCCTTGGCTACTTTT | 4485 |
| Db | 4321 | GTTTCAGTCAACTAGTGCACCAATTCGACATGATGAATAGACTGACACTATCCTTGGCTACTTTT | 4380 |
| Qy | 4486 | CAAAATTTACAAAATTTACCACTCAACTTCGTGCTCATAAAATGCAAAAGAAATGGGAATCCGAA | 4545 |
| Db | 4381 | CAAAATTTACAAAATTTACCACTCAACTTCGTGCTCATAAAATGCAAAAGAAATGGGAATCCGAA | 4440 |
| Qy | 4546 | GAATGTGTTGAAATATCAAAAAATTTCAATACCTCTATGAAGCTCACTTTAACTGAAGAGACA   | 4605 |
| Db | 4441 | GAATGTGTTGAAATATCAAAAAATTTCAATACCTCTATGAAGCTCACTTTAACTGAAGAGACA   | 4500 |
| Qy | 4606 | GTATATGGATTAAGAAATTTAGATGTGTATAGATTTCCAAGTGGCTGCAGCACCAATGCT      | 4665 |
| Db | 4501 | GTATATGGATTAAGAAATTTAGATGTGTATAGATTTCCAAGTGGCTGCAGCACCAATGCT      | 4560 |
| Qy | 4666 | GGCTATGGCAATGCTTCAAACTGATTTCTACAAAACCTCGCTGCGCCCTCCAGATGGT        | 4725 |
| Db | 4561 | GGCTATGGCAATGCTTCAAACTGATTTCTACAAAACCTCGCTGCGCCCTCCAGATGGT        | 4620 |
| Qy | 4726 | CCTCTCTGAAATGTTCTATGTAGTACCAACATCACCTTTTATAGCATCAGCATTAAGCTGGAGT  | 4785 |
| Db | 4621 | CCTCTCTGAAATGTTCTATGTAGTACCAACATCACCTTTTATAGCATCAGCATTAAGCTGGAGT  | 4680 |
| Qy | 4786 | GAACCTGCTGTCATTACTGACCAACCAATGTTTATCTGATTTGATGTCAATCGGTAGATAAT    | 4845 |
| Db | 4681 | GAACCTGCTGTCATTACTGACCAACCAATGTTTATCTGATTTGATGTCAATCGGTAGATAAT    | 4740 |

|    |      |  |      |
|----|------|--|------|
| Qy | 4846 | GATGAAATTTAATAATATATCCCTTCATCAAGTCAAATGAAGAAAATAAACCATAGAAATTAAA   | 4901 |
| Db | 4741 | GATGAAATTTAATAATATCCCTTCATCAAGTCAAATGAAGAAAATAAACCATAGAAATTTAAA    | 4800 |
| Qy | 4906 | GATTTAGAAATATTTACAAGGTATTCGTAGTGATCACTGCATTTACTCTGGGAACATTAGT      | 4965 |
| Db | 4801 | GATTTAGAAATATTTACAAGGTATTCGTAGTGATCACTGCATTTACTCTGGGAACATTAGT      | 4860 |
| Qy | 4966 | GCTGCATATGTAGAAGGAAAGTCAAGTGTGGAATGATTTACTTACTTTAGAAATCAGCC        | 5025 |
| Db | 4861 | GCTGCATATGTAGAAGGAAAGTCAAGTGTGGAATGATTTACTTACTTTAGAAATCAGCC        | 4920 |
| Qy | 5026 | CCAAGAACCCACTTAACAACATGACATTTTCAGAAATACAGATGAGATTAACAANAATTT       | 5085 |
| Db | 4921 | CCAAGAACCCACTTAACAACATGACATTTTCAGAAATACAGATGAGATTAACAANAATTT       | 4980 |
| Qy | 5086 | CAATTAACGTTTTCTTCTCTCTCTCAACCTAAATGGAATATCAAGTATATCAAGCTCTG        | 5145 |
| Db | 4981 | CAATTAACGTTTTCTTCTCTCTCTCAACCTAAATGGAATATCAAGTATATCAAGCTCTG        | 5040 |
| Qy | 5146 | GTTTACCAGAGAAGATGATCCTACTGCTGTCAGATTCCAACTTCAACCTCAGTATTTATACAGAAA | 5205 |
| Db | 5041 | GTTTACCAGAGAAGATGATCCTACTGCTGTCAGATTCCAACTTCAACCTCAGTATTTATACAGAAA | 5100 |
| Qy | 5206 | ACCAACACATTCGTTCATTTGCCAATCTTAGAAGGACTAAAGGTGGACATACATACAAATATC    | 5265 |
| Db | 5101 | ACCAACACATTCGTTCATTTGCCAATCTTAGAAGGACTAAAGGTGGACATACATACAAATATC    | 5160 |
| Qy | 5266 | AGTGTTTTACCAGTCAATAGTGTGCTGGTGAGGTCCTAAAGGTCCCGATGAGAAATAACCATG    | 5325 |
| Db | 5161 | AGTGTTTTACCAGTCAATAGTGTGCTGGTGAGGTCCTAAAGGTCCCGATGAGAAATAACCATG    | 5220 |
| Qy | 5326 | GATATCAAAGCTCCAGCACGACCAACAAACCAAACCCCTATTTATGATGCCACAGGA          | 5385 |
| Db | 5221 | GATATCAAAGCTCCAGCACGACCAACAAACCAAACCCCTATTTATGATGCCACAGGA          | 5280 |
| Qy | 5386 | AAACTGCTTGTGACTTCAACAAATTAACATCAGAAATGCCAATATGTTACTACAGTGAT        | 5445 |
| Db | 5281 | AAACTGCTTGTGACTTCAACAAATTAACATCAGAAATGCCAATATGTTACTACAGTGAT        | 5340 |
| Qy | 5446 | GATCATGGACCAATAAAAATGTACAGTCTTTGCGACAGAAACAGAGCTCAGCATGAT          | 5505 |
| Db | 5341 | GATCATGGACCAATAAAAATGTACAGTCTTTGCGACAGAAACAGAGCTCAGCATGAT          | 5400 |
| Qy | 5506 | GGAATGTAAACAAAGTGTTATGTCATATTTTAAATAAGCAAGGCCCATATTTTACAAT         | 5565 |
| Db | 5401 | GGAATGTAAACAAAGTGTTATGTCATATTTTAAATAAGCAAGGCCCATATTTTACAAT         | 5460 |
| Qy | 5566 | GAAGGCTTTCCTAACCCCTCATGTACAGAAAGAAAGCAAAAGTTTATGTCGAATGAAGAA       | 5625 |
| Db | 5461 | GAAGGCTTTCCTAACCCCTCATGTACAGAAAGAAAGCAAAAGTTTATGTCGAATGAAGAA       | 5520 |
| Qy | 5626 | ATCTACATCATAGGTGCTGTAATGCAATGATTCCTGGCAATGAAGCAAAAATTTGC           | 5685 |
| Db | 5521 | ATCTACATCATAGGTGCTGTAATGCAATGATTCCTGGCAATGAAGCAAAAATTTGC           | 5580 |
| Qy | 5686 | AATGGACCACTGAAACCAAAAAGCAATCTATTTAAATTTAGAGCTACAAAATATTATG         | 5745 |
| Db | 5581 | AATGGACCACTGAAACCAAAAAGCAATCTATTTAAATTTAGAGCTACAAAATATTATG         | 5640 |
| Qy | 5746 | GGAACAATTTTACTGACTCTGATTAATCTGACCCCTGTTAAGACTTTTAGGGGAAGCACTTCA    | 5805 |
| Db | 5641 | GGAACAATTTTACTGACTCTGATTAATCTGACCCCTGTTAAGACTTTTAGGGGAAGCACTTCA    | 5700 |
| Qy | 5806 | GAAAGAACCGTAGAGATCATTTCTTCGTCACATTTGTGTATTCCTTTCAATAATTCCTCTT      | 5865 |
| Db | 5701 | GAAAGAACCGTAGAGATCATTTCTTCGTCACATTTGTGTATTCCTTTCAATAATTCCTCTT      | 5760 |
| Qy | 5866 | GGAACGACTATTTTGTCAATTTGCAAAATTCGACAGAGCAGAAAGAGGTGGCACANATC        | 5925 |
| Db | 5761 | GGAACGACTATTTTGTCAATTTGCAAAATTCGACAGAGCAGAAAGAGGTGGCACANATC        | 5820 |
| Qy | 5926 | TCTCCTCAGAGATGCAGAAATTTATTTGACACTAAATTTGAAGCTGGATCAGCTCATCAGATG    | 5985 |

```
Db      5821  ||||| TCCTCAGATGCGAATATTGACACTAAATTGAAGCTGGATCAGCTCATCAGTG 5880
Qy      5986  GCAGACTGGAACCTGAAGACGAGAGATTAAACGCG-----
Db      5881  GCAGACTGGAACCTGAAGACGAGAGATTAAACGCGGTACTTAGTTATAGAAAATCCATC 5940
Qy      6021  --GCCAATAGCAAGAAATCTTCCTGCAACATGTTGAGAGCTTTGCAACACACAC 6078
Db      5941  AAGCCAATAGCAAGAAATCTTCCTGCAACATGTTGAGAGCTTTGCAACACACAC 6000
Qy      6079  CTAAGGTTCAAGAAATTTTCGGAATACCAAAATTTCTTCAGGATCTTTCTCAACT 6138
Db      6001  CTAAGGTTCAAGAAATTTTCGGAATACCAAAATTTCTTCAGGATCTTTCTCAACT 6060
Qy      6139  GATGCTGATCTGCTTTGGAATAGAGCAAAACCGTTTCCCAACATAAACAATATAAT 6198
Db      6061  GATGCTGATCTGCTTTGGAATAGAGCAAAACCGCTTCCCAACATAAACAATATAAT 6117
Qy      6199  AATAATAACAGATTAAGCTGATAGCTGACGCTAGTGTTCAGGTTCCGATTTATTAAT 6258
Db      6118  AATAATAACAGATTAAGCTGATAGCTGACGCTAGTGTTCAGGTTCCGATTTATTAAT 6177
Qy      6259  GCCAGCTATATTCTCGTTATTTATGTCCTCAAAATGAATTTATGCTACTCAAGTCCACTA 6318
Db      6178  GCCAGCTATATTCTCGTTATTTATGTCCTCAAAATGAATTTATGCTACTCAAGTCCACTA 6237
Qy      6319  CCAGGAACAGTTGGAGATTTTGGAGAAATGGTGTGGGAAACCGAGGCAAAACATTAATA 6378
Db      6238  CCAGGAACAGTTGGAGATTTTGGAGAAATGGTGTGGGAAACCGAGGCAAAACATTAATA 6297
Qy      6379  ATGCTAACACAGTGTGTTGAAAAGGACGATCAGATGCCATCAGTATTTGGCCAGAGAC 6438
Db      6298  ATGCTAACACAGTGTGTTGAAAAGGACGATCAGATGCCATCAGTATTTGGCCAGAGAC 6357
Qy      6439  AACAAAGCAGTTACTGCTTTGGAGATATAGTGATTAACAAAGCTAATGAGGATGTTCAA 6498
Db      6358  AACAAAGCAGTTACTGCTTTGGAGATATAGTGATTAACAAAGCTAATGAGGATGTTCAA 6417
Qy      6499  ATAGATGGAATATACAGGATCTGAAAATTTGAAAGGATGAGGATGAGATGCTGTTGGA 6558
Db      6418  ATAGATGGAATATACAGGATCTGAAAATTTGAAAGGATGAGGATGAGATGCTGTTGGA 6477
Qy      6559  CAGGTAACTTTACTGCTGCGGACGAGCATGCGGTTCCTGGAACAGCGCCCTCTAAT 6618
Db      6478  CAGGTAACTTTACTGCTGCGGACGAGCATGCGGTTCCTGGAACAGCGCCCTCTAAT 6537
Qy      6619  CACTTTGTGAAGTTGGTTCGAGCAAGCAGGACATGACACACACCTATGATTTGTTCA 6678
Db      6538  CACTTTGTGAAGTTGGTTCGAGCAAGCAGGACATGACACACACCTATGATTTGTTCA 6597
Qy      6679  TGCAAGTGGAGTTGGAGAACTGAGATTTTATGCTGTGGAACATTTAAACACAAAT 6738
Db      6598  TGCAAGTGGAGTTGGAGAACTGAGATTTTATGCTGTGGAACATTTAAACACAAAT 6657
Qy      6739  ATAAATGACCAATGTTGGATATATAGGATAGTAGTGAATCTGAGAGTGAAGA 6798
Db      6658  ATAAATGACCAATGTTGGATATATAGGATAGTAGTGAATCTGAGAGTGAAGA 6717
Qy      6799  ATGTCATGTCGAGAACTGCGACAGTATATCTTTTACACAGTGCATTTCTGATCTC 6858
Db      6718  ATGTCATGTCGAGAACTGCGACAGTATATCTTTTACACAGTGCATTTCTGATCTC 6777
Qy      6859  TTAATAAATAGGGAAGTAAATCAGCCCATCTGTTTGTGTTAACTAATTCAGCACTTCAGAAG 6918
Db      6778  TTAATAAATAGGGAAGTAAATCAGCCCATCTGTTTGTGTTAACTAATTCAGCACTTCAGAAG 6837
Qy      6919  ATGCACTTTGCAAGCCCATGGAAGTGTGTTGAGCTTGAATGGAAGAAACCACTATG 6978
Db      6838  ATGCACTTTGCAAGCCCATGGAAGTGTGTTGAGCTTGAATGGAAGAAACCACTATG 6897
```

```
AX644236
LOCUS      AX644236              7225 bp      DNA      linear      PAT 27-FEB-2003
DEFINITION Sequence 10 from Patent WO02090530.
ACCESSION  AX644236
VERSION    AX644236.1      GI:28610295
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1
AUTHORS   Lee, E.A., Walla, N.K., Baughn, M.R., Ison, C.H., Gururajan, R.,
            Arvizu, C., Yao, M.G., Jackson, J.L., Tang, T.Y., Yue, H., Tran, B.,
            Ding, L., Lu, D.A., Lal, P.G. and Warren, B.A.
TITLE     Kinases and phosphatases
JOURNAL   Patent: WO 02090530-A 10 14-NOV-2002;
            Incyte Genomics, Inc. (US)
FEATURES   Location/Qualifiers
            1..7225
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            /note="Incyte ID No: 7480588CB1"

ORIGIN
Query Match      95.7%; Score 6799.8; DB 6; Length 7225;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 6962; Conservative 0; Mismatches 17; Indels 72; Gaps 8;

Qy      106  ATGGATTTTCTTATCATTTTCTTTTACTTTTATTTGGACTTCAGAGACAGGTTGAT 165
Db      1    ATGGATTTTCTTATCATTTTCTTTTACTTTTATTTGGACTTCAGAGACAGGTTGAT 60
Qy      166  GTTTCCAATGTCGTTCTGTTACTAGGTACGATATAACCATCTCTTCAATTTCTACAACA 225
Db      61  GTTTCCAATGTCGTTCTGTTACTAGGTACGATATAACCATCTCTTCAATTTCTACAACA 120
Qy      226  TACACTCACCTGTTACTAGATAGTACACCAAAATGTAACAAAACAGGCGCTCCAGTC 285
Db      121  TACACTCACCTGTTACTAGATAGTACACCAAAATGTAACAAAACAGGCGCTCCAGTC 165
Qy      286  TTCTAGCGCGGGAAGAGTCGATCTGCTGGGATTTCTTCTGTTGGAATACACACCT 345
Db      166  TTCTAGCGCGGGAAGAGTCGATCTGCTGGGATTTCTTCTGTTGGAATACACACCT 225
Qy      346  AATCCAATGGAAGATTATATCTTACATTGTCAAATATAAGAAAGTTGTCCGTGGATG 405
Db      226  AATCCAATGGAAGATTATATCTTACATTGTCAAATATAAGAAAGTTGTCCGTGGATG 285
Qy      406  CAAACAGTATATACAAAGTCAGATCAAAAGCCAGACAGTCTGGAAGTTCTTCTTACTAAT 465
Db      286  CAAACAGTATATACAAAGTCAGATCAAAAGCCAGACAGTCTGGAAGTTCTTCTTACTAAT 345
Qy      466  CTTAATCTCGGAACAAACATATGAAATTAAGTTGCTGCTGAAAACAGTCTGCGCATGGA 525
Db      346  CTTAATCTCGGAACAAACATATGAAATTAAGTTGCTGCTGAAAACAGTCTGCGCATGGA 405
Qy      526  GTGTTTAGTGATCAATTTCTTCCAACTCGAGAAAGTGTCTCCAGGAAAGTGTGTAAT 585
Db      406  GTGTTTAGTGATCAATTTCTTCCAACTCGAGAAAGTGTCTCCAGGAAAGTGTGTAAT 465
Qy      586  CTCACAGTTGAGGCTTACAAACGCTTTCAGCAGTTAAGCTGATTTGTTACTCTGGCAA 645
Db      466  CTCACAGTTGAGGCTTACAAACGCTTTCAGCAGTTAAGCTGATTTGTTACTCTGGCAA 525
Qy      646  CCAAAATGGCAAAATTAACAGCTTCAAGATTAGTGTCAAAATGCGCAAGTGGGATGTA 705
Db      526  CCAAAATGGCAAAATTAACAGCTTCAAGATTAGTGTCAAGATTAGTGTGGGATGTA 585
Qy      706  GTGAAAGATGTTCAATCAGAGTAGGACATTTTGAATGAGGAAATTCAGAAATGCAAT 765
Db      586  GTGAAAGATGTTCAATCAGAGTAGGACATTTTGAATGAGGAAATTCAGAAATGCAAT 645
```

|    |      |   |  |                      |      |
|----|------|---|--|----------------------|------|
| Qy | 766  | GAGATAGTGAATCTTTTATGAGTAGTACAGCGAGCCCTTCTCCAA   | CCCTTGGTAGAGTT                                     | 825                  |      |
| Db | 646  | GAGAAATAGTGAATCTTTTATGAGTAGTACAGCGAGCCCTTCTCCAA | CCCTTGGTAGAGTT                                     | 705                  |      |
| Qy | 826  | ACACTCCATCGCGTACCAACATTCATCAAGCAGCTTGA          | CACAGATGAGATCAGCTCT                                | 885                  |      |
| Db | 706  | ACACTCCATCGCGTACCAACATTCATCAAGCAGCTTGA          | CACAGATGAGATCAGCTCT                                | 765                  |      |
| Qy | 886  | GTGTGGAAGAGCCCTTACAGTTTGTAGTGA                  | CACACTTGGAGCCTTATACAACTATCTTT                      | 945                  |      |
| Db | 766  | GTGTGGAAGAGCCCTTACAGTTTGTAGTGA                  | CACACTTGGAGCCTTATACAACTATCTTT                      | 825                  |      |
| Qy | 946  | TTTGAAGTTTCAGCTGCTA                             | CAACTGAGCAGGTTATTTGATAGTACGATTTGT                  | 1005                 |      |
| Db | 826  | TTTGAAGTTTCAGCTGCTA                             | CAACTGAGCAGGTTATTTGATAGTACGATTTGT                  | 885                  |      |
| Qy | 1006 | CCAGAACTCAGTCCGTGAAGGACACCA                     | CAAAACCTGCGTAA                                     | CAGGCAACATCACAAGGAAG | 1065 |
| Db | 886  | CCAGAACTCAGTCCGTGAAGGACACCA                     | CAAAACCTGCGTAA                                     | CAGGCAACATCACAAGGAAG | 945  |
| Qy | 1066 | TCCTTTTCAATTTTATGGGACCCCAACTATAGTAA             | CAGGGAATTTAGTTATAGAGTT                             | 1125                 |      |
| Db | 946  | TCCTTTTCAATTTTATGGGACCCCAACTATAGTAA             | CAGGGAATTTAGTTATAGAGTT                             | 1005                 |      |
| Qy | 1126 | GAAATATATGGAACCAT-----                          | CAGGTGCGAATTTTGGATTA                               | CACGACAAAGACCTCAAG   | 1179 |
| Db | 1006 | GAAATATATGGAACCATCAGGTGCGAGT                    | CGCATTTTGGATTA                                     | CACGACAAAGACCTCAAG   | 1065 |
| Qy | 1180 | TTTGCATTCCTAACCTAACACCATTTACATGATGATGTC         | ATATATTTGGGCTGGAACC                                | 1239                 |      |
| Db | 1066 | TTTGCATTCCTAACCTAACACCATTTACATGATGATGTC         | ATATATTTGGGCTGGAACC                                | 1125                 |      |
| Qy | 1240 | AGTGCAGGAGCTGGGCCCAAGTCAAAATATTTTCAGTAT         | TTTCAGTATTCCTCCACAGATTTTCCAGGG                     | 1299                 |      |
| Db | 1126 | AGTGCAGGAGCTGGGCCCAAGTCAAAATATTTTCAGTAT         | TTTCAGTATTCCTCCACAGATTTTCCAGGG                     | 1185                 |      |
| Qy | 1300 | GCAGTGTGTTGATTTACAACTTCGACAGGTAGAA              | TCCACGCAAGTAAGAAATTTACTTGGAG                       | 1359                 |      |
| Db | 1186 | GCAGTGTGTTGATTTACAACTTCGACAGGTAGAA              | TCCACGCAAGTAAGAAATTTACTTGGAG                       | 1245                 |      |
| Qy | 1360 | AAACCGACGACAAACAAATGGAAATTTAA                   | CCAAATACCGAGTGAAGTCTAGTTC                          | CACAG                | 1419 |
| Db | 1246 | AAACCGACGACAAACAAATGGAAATTTAA                   | CCAAATACCGAGTGAAGTCTAGTTC                          | CACAG                | 1305 |
| Qy | 1420 | ACAGGAATAATTTTGGAAATAC                          | TTTGTCTCACTCGAAATATAGTATATAAAATGACCCC              | 1479                 |      |
| Db | 1306 | ACAGGAATAATTTTGGAAATAC                          | TTTGTCTCACTCGGAATATAGTATATAAAATGACCCC              | 1362                 |      |
| Qy | 1480 | ATGCTCCAGAAATTTGGAACATAGTAGAGCCAA               | TGGTAGTTATATGAGGGTTCAGCA                           | 1539                 |      |
| Db | 1363 | ATGCTCCAGAAATTTGGAACATAGTAGTAGCCAA              | TGGTAGTTATATGAGGGTTCAGCA                           | 1422                 |      |
| Qy | 1540 | GAGATGTCCTGACCTTCACTCACTTGTCTACATTTATATATA      | TAACAGCCATCCAGATAAA                                | 1599                 |      |
| Db | 1423 | GAGATGTCCTGACCTTCACTCACTTGTCTACATTTATATA        | TATACAGCCATCCAGATAAA                               | 1482                 |      |
| Qy | 1600 | AACCTTCTCGCAAGGAATAGAGCTGGAAGAC                 | CAGACTTCCACAGTTGTAACTACAAGGAAT                     | 1659                 |      |
| Db | 1483 | AACCTTCTCGCAAGGAATAGAGCTGGAAGAC                 | CAGACTTCCACAGTTGTAACTACAAGGAAT                     | 1542                 |      |
| Qy | 1660 | CAGTATATTA                                      | CTGACATTTGCAAGCTGTCTTATGTTATCAGGAGACTTGTAACCT      | 1719                 |      |
| Db | 1543 | CAGTATATTA                                      | CTGACATTTGCAAGCTGTCTTATGTTATCAGGAGACTTGTAACCT      | 1602                 |      |
| Qy | 1720 | TTCACTGAGCA                                     | CATGATAGTGTATCTGCTTCAACATCATGCGGAAGGACCA           | 1779                 |      |
| Db | 1603 | TTCACTGAGCA                                     | CATGATAGTGTATCTGCTTCAACATCATGCGGAAGGACCA           | 1662                 |      |
| Qy | 1780 | GTTCTCAGTGT                                     | TAGGACACGTCCAGCAAGTCCCAAGCTCCATTTAAATTTATATACTATAA | 1839                 |      |
| Db | 1663 | GTTCTCAGTGT                                     | TAGGACACGTCCAGCAAGTCCCAAGCTCCATTTAAATTTATATACTATAA | 1722                 |      |
| Qy | 1840 | AAATATTAGTTCCTTCATCTATTTTGTGTATATTTGGGAT        | CCTCCAGAAATATCCCAATGGAAAA                          | 1899                 |      |

|    |      |  |  |      |  |
|----|------|--|--|------|--|
| DB | 1723 |  | AAATATTAGTTCCTTCATCTATTTTGTGTATATTTGGATCCTCAGAATATCCCAATGGAAA    | 1782 |  |
| QY | 1900 |  | ATAACTCACTATACGATTTATGCAATGGAATTTGGATACAAACAGAGCAATTTCCAGATAACT  | 1959 |  |
| DB | 1783 |  | ATAACTCACTATACGATTTATGCAATGGAATTTGGATACAAACAGAGCAATTTCCAGATAACT  | 1842 |  |
| QY | 1960 |  | ACCATAGATAACAGCTTTCTCATTAACAG-----GGTTAAAGAAATACACAAA            | 2007 |  |
| DB | 1843 |  | ACCATAGATAACAGCTTTCTCATTAACAGGTAGAAAACAAATGGTTAAAGAAATACACAAA    | 1902 |  |
| QY | 2008 |  | TACAAAATGAGAGTGGCAGCCTCAACCCACGATGGAGAAAGTCTTTGTCTGAAGAAAT       | 2067 |  |
| DB | 1903 |  | TACAAAATGAGAGTGGCAGCCTCAACCCACGTTGGAGAAAGTCTTTGTCTGAAGAAAT       | 1962 |  |
| QY | 2068 |  | GACATCTTTGTGAGAACTTCAGAAAGATGAACCGGAATCATCACCTCAAGATGTCGAAGTA    | 2127 |  |
| DB | 1963 |  | GACATCTTTGTGAGAACTTCAGAAAGATGAACCGGAATCATCACCTCAAGATGTCGAAGTA    | 2022 |  |
| QY | 2128 |  | ATTGATGTTACCGCAGATGAATTAAGGTTGAAGTGGTCACACCCGAAAAAGCCCAATGGG     | 2187 |  |
| DB | 2023 |  | ATTGATGTTACCGCAGATGAATTAAGGTTGAAGTGGTCACACCCGAAAAAGCCCAATGGG     | 2082 |  |
| QY | 2188 |  | ATCATTTATGCTTATGAAGTGCTATATAAAAAATATAGATACTTTTATATATGAAGAACACA   | 2247 |  |
| DB | 2083 |  | ATCATTTATGCTTATGAAGTGCTATATAAAAAATATAGATACTTTTATATATGAAGAACACA   | 2142 |  |
| QY | 2248 |  | TCAACACACAGACATAATATTAAAGGAACCTTAAGACCTCACACCTCTATAACATTTCTGTGA  | 2307 |  |
| DB | 2143 |  | TCAACACACAGACATAATATTAAAGGAACCTTAAGACCTCACACCTCTATAACATTTCTGTGA  | 2202 |  |
| QY | 2308 |  | AGGTCTTACACACAGATTTGGTCATGGCAATCAGGTATCTTTTACTCTCTGTAAGGACT      | 2367 |  |
| DB | 2203 |  | AGGTCTTACACACAGATTTGGTCATGGCAATCAGGTATCTTTTACTCTCTGTAAGGACT      | 2262 |  |
| QY | 2368 |  | TCGAGAGACTGTCCTGATAGTGCACAGAAAAATATCATTTACAAAAATATTTCTTCTGGGA    | 2427 |  |
| DB | 2263 |  | TCGAGAGACTGTCCTGATAGTGCACAGAAAAATATCATTTACAAAAATATTTCTTCTGGGA    | 2322 |  |
| QY | 2428 |  | GAGATTGAGCTATCATTTCTTTCCCCCAAGTAGTCCCAATGGAAATCATAAAAAATATACA    | 2487 |  |
| DB | 2323 |  | GAGATTGAGCTATCATTTCTTTCCCCCAAGTAGTCCCAATGGAAATCATAAAAAATATACA    | 2382 |  |
| QY | 2488 |  | ATTTATCTCAAGAGAAGTAATGGAAATGAGGAAGAACTATAAATACAACCTCTTTAAACC     | 2547 |  |
| DB | 2383 |  | ATTTATCTCAAGAGAAGTAATGGAAATGAGGAAGAACTATAAATACAACCTCTTTAAACC     | 2442 |  |
| QY | 2548 |  | CAAAAACATTAAAGTACTGAAGAAATATACCCAATATATCATTTGAGGTGTCGTAGTACA     | 2607 |  |
| DB | 2443 |  | CAAAAACATTAAAGGACTGAAGAAATATACCCAATATATCATTTGAGGTGTCGTAGTACA     | 2502 |  |
| QY | 2608 |  | CTGAAGGTTGAAGAGTTCGGAGTGCTCCCATTAAGTATACTGACGGAGGAAGATGCTCCT     | 2667 |  |
| DB | 2503 |  | CTGAAGGTTGAAGAGTTCGGAGTGCTCCCATTAAGTATACTGACGGAGGAAGATGCTCCT     | 2562 |  |
| QY | 2668 |  | GAATTTCTCCCTCAAGACTTCTCTGTAAAAACAGTTGTCTGGTGTACGGTGAAGTTGTCA     | 2727 |  |
| DB | 2563 |  | GAATTTCTCCCTCAAGACTTCTCTGTAAAAACAGTTGTCTGGTGTACGGTGAAGTTGTCA     | 2622 |  |
| QY | 2728 |  | TGGCAACACCCCTGGAGCCAAATGGAATTTATCTTTTATTTACACAGTTTATGTCT---GG    | 2784 |  |
| DB | 2623 |  | TGGCAACACCCCTGGAGCCAAATGGAATTTATCTTTTATTTACACAGTTTATGTCTGGAGG    | 2682 |  |
| QY | 2785 |  | AATAGATCATCATTTAAAACTATTAAATGTCACCTGAAACATCATTTGGAGTTTATCAGATTTG | 2844 |  |
| DB | 2683 |  | AATAGATCATCATTTAAAACTATTAAATGTCACCTGAAACATCATTTGGAGTTTATCAGATTTG | 2742 |  |
| QY | 2845 |  | GATTATAATGTTGAATACAGTGCTTATGTAACAGCTAGCACAGATTTGGTGTATGGGAAA     | 2904 |  |
| DB | 2743 |  | GATTATAATGTTGAATACAGTGCTTATGTAACAGCTAGCACAGATTTGGTGTATGGGAAA     | 2802 |  |
| QY | 2905 |  | ACAGGAAGCAATATCATTTAGCTTTCAAACACAGAGGGAGCACCAACGCGATCTCCCAA      | 2964 |  |





QY 5125 ATCCAAGTATATCAAGCTCTGGTTTACCGAGAAGATGATCCTACTGCTGTCCAGATTCCAC 5184  
DB 5020 ATCCAAGTATATCAAGCTCTGGTTTACCGAGAAGATGATCCTACTGCTGTCCAGATTCCAC 5079  
QY 5185 AACCTCAGTATATACAGAAACCAACACATTTCTGTCATTTGCAATGCTAGAGGACTAAAA 5244  
DB 5080 AACCTCAGTATATACAGAAACCAACACATTTCTGTCATTTGCAATGCTAGAGGACTAAAA 5139  
QY 5245 GGTGGACATACATCAATATACAGTTTACGAGTCAATAGTGTGCTGTGAGGTCCTCAAG 5304  
DB 5140 GGTGGACATACATCAATATACAGTTTACGAGTCAATAGTGTGCTGTGAGGTCCTCAAG 5199  
QY 5305 GTTCCGATGAGATTAACCATGGATATCAAGCTCCAGCAGCAACCAAAACCAACCAACC 5364  
DB 5200 GTTCCGATGAGATTAACCATGGATATCAAGCTCCAGCAGCAACCAAAACCAACCAACC 5259  
QY 5365 CCTATTATGATGCCACAGGAAACTCTTGTGACTTCAACCAATTAACAATCAGAAATG 5424  
DB 5260 CCTATTATGATGCCACAGGAAACTCTTGTGACTTCAACCAATTAACAATCAGAAATG 5319  
QY 5425 CCAATATGTTACTACAGTGATGATCATGACCAATAAAAATGTACAAAGTGTGCGACA 5484  
DB 5320 CCAATATGTTACTACAGTGATGATCATGACCAATAAAAATGTACAAAGTGTGCGACA 5379  
QY 5485 GAAACAGAGCTCAGCATGATGAAATGTAAACAAAGTGTGATGATGATGATGATGATGAT 5544  
DB 5380 GAAACAGAGCTCAGCATGATGAAATGTAAACAAAGTGTGATGATGATGATGATGATGAT 5439  
QY 5545 GCAAGGCCATATTTTACAAATGAAGCTTTTCTTAACCTCCATGTCAGAGCAAGCAAGACA 5604  
DB 5440 GCAAGGCCATATTTTACAAATGAAGCTTTTCTTAACCTCCATGTCAGAGCAAGCAAGACA 5499  
QY 5605 AAGTTTGTAGTGGCAATGAAGAAATCTCATCATAGTGTGCTGATTAATGATGATGATGATGAT 5664  
DB 5500 AAGTTTGTAGTGGCAATGAAGAAATCTCATCATAGTGTGCTGATTAATGATGATGATGATGAT 5559  
QY 5665 GGCATGAAGACAAAATTTGCAATGGACCACTGAAACCAAAAAGCAATATTTTAA 5724  
DB 5560 GGCATGAAGACAAAATTTGCAATGGACCACTGAAACCAAAAAGCAATATTTTAA 5619  
QY 5725 TTTAGAGTCAAAATATTTATGGACAAATTTACTGACTCTGATTAATTTCTGACCTGTTAAG 5784  
DB 5620 TTTAGAGTCAAAATATTTATGGACAAATTTACTGACTCTGATTAATTTCTGACCTGTTAAG 5679  
QY 5785 ACTTTAGGGGAAGGACTTTTCAAGAAACCGTAGAGATCAATTTCTTCCTGCTCACTTTGTGT 5844  
DB 5680 ACTTTAGGGGAAGGACTTTTCAAGAAACCGTAGAGATCAATTTCTTCCTGCTCACTTTGTGT 5739  
QY 5845 ATCTTTCAATAATTTCTCTTGGAAACGCTATTTTGGCAATTTGCAATTTGCAATTTGCAAG 5904  
DB 5740 ATCTTTCAATAATTTCTCTTGGAAACGCTATTTTGGCAATTTGCAATTTGCAAG 5799  
QY 5905 CAGAAAGAGTGGCATACTCTCTCAGGATGCAAGAAATTTTCAATTAATTTGCAATTTGCAAG 5964  
DB 5800 CAGAAAGAGTGGCATACTCTCTCAGGATGCAAGAAATTTTCAATTAATTTGCAATTTGCAAG 5859  
QY 5965 CTGGATCAGCTCATCAGTGGCAGACCTGGAATGAAAGCAGAGATTTAAACGCG --- 6020  
DB 5860 CTGGATCAGCTCATCAGTGGCAGACCTGGAATGAAAGCAGAGATTTAAACGCGTTA 5919  
QY 6021 -----GCCAATAAGCAAGAAATCTCTCTGCAACATGTTGAA 6057  
DB 5920 CTTTAGTTATAGAAAATCCATCAAGCCCAATAAGCAAGAAATCTCTCTGCAACATGTTGAA 5979  
QY 6058 GAGCTTTGCAAAAACCAACCTTAAAGTTTCAAGAAAGAAATTTTCGGAATTTTCAAAATTT 6117  
DB 5980 GAGCTTTGCAAAAACCAACCTTAAAGTTTCAAGAAAGAAATTTTCGGAATTTTCAAAATTT 6039  
QY 6118 CTTTCAGATCTTTCTTCACTGATGCTGCTGCTTGGATAGACAAACCAACCTTTC 6177  
DB 6040 CTTTCAGATCTTTCTTCACTGATGCTGCTTGGATAGACAAACCAACCTTTC 6099  
QY 6178 CCAACATRAAACCATATAATATAACAGAGTAAAGCTGATAGCTGACGCTAGTGT 6237

DB 6100 CCAACATRAAACCATATAATATAACAGAGTAAAGCTGATAGCTGACGCTAGTGT 6156  
QY 6238 CCAAGTTTCGATATATTAATGTCAGCTATATTTCTGTTTATTTATGTCCTCAAAATGAATTT 6297  
DB 6157 CCAAGTTTCGATATATTAATGTCAGCTATATTTCTGTTTATTTATGTCCTCAAAATGAATTT 6216  
QY 6298 ATTGCTACTCAAGGTCACCTACAGGAAACGTTGGAGATTTTGGAGAAATGTTGGAGAA 6357  
DB 6217 ATTGCTACTCAAGGTCACCTACAGGAAACGTTGGAGATTTTGGAGAAATGTTGGAGAA 6276  
QY 6358 ACCAGGGCAAAAACATTTAGTAAATGCTAAACACAGTGTGTTGAAAAGGACGATCAGATGC 6417  
DB 6277 ACCAGAGCAAAAACATTTAGTAAATGCTAAACACAGTGTGTTGAAAAGGACGATCAGATGC 6336  
QY 6418 CATCAGTATTTGGCCAGAGGACCAACAGCCAGTTTACTGTCTTTGGAGATATAGTGAATTACA 6477  
DB 6337 CATCAGTATTTGGCCAGAGGACCAACAGCCAGTTTACTGTCTTTGGAGATATAGTGAATTACA 6396  
QY 6478 AAGCTAATGGAGATGTTTCAAAATAGATGGAATCATCAGGATCTGAAAATTTGAAAGGCAT 6537  
DB 6397 AAGCTAATGGAGATGTTTCAAAATAGATGGAATCATCAGGATCTGAAAATTTGAAAGGCAT 6456  
QY 6538 GGGGATTCGATGACTGTTTGCAGAGTGTAACTTTTACTGCTGCCAGAGCATGGGTTCT 6597  
DB 6457 GGGGATTCGATGACTGTTTGCAGAGTGTAACTTTTACTGCTGCCAGAGCATGGGTTCT 6516  
QY 6598 GAGAACAGCGCCCTCTAAATTCATTTGTGAGTGTGTTTCGAGCAAGCAGGCGACATGAC 6657  
DB 6517 GAGAACAGCGCCCTCTAAATTCATTTGTGAGTGTGTTTCGAGCAAGCAGGCGACATGAC 6576  
QY 6658 ACCACACTATGATGTTTCACTGCTGAGTGTGGAAGAACTGGAGTTTATTTGCT 6717  
DB 6577 ACCACACTATGATGTTTCACTGCTGAGTGTGGAAGAACTGGAGTTTATTTGCT 6636  
QY 6718 CTGGACCATTTAAACCAACATATAATGACCATGATTTTGTGGATATATATGACCTAGTA 6777  
DB 6637 CTGGACCATTTAAACCAACATATAATGACCATGATTTTGTGGATATATATGACCTAGTA 6696  
QY 6778 GCTGAATCAGAAAGTGAAGAAATGTCATGTTGCAGAACTGTCACAGATATATCTTTTAA 6837  
DB 6697 GCTGAATCAGAAAGTGAAGAAATGTCATGTTGCAGAACTGTCACAGATATATCTTTTAA 6756  
QY 6838 CACCACTGATTTCTGGATCTCTTATCAAAATAGGGAAGTAATCAGCCCATCTGTTTGT 6897  
DB 6757 CACCACTGATTTCTGGATCTCTTATCAAAATAGGGAAGTAATCAGCCCATCTGTTTGT 6816  
QY 6898 AACTATTCAGCATTTTCAAGAGTGGACTCTTTGGAGCCCATGGAAGGTGATGTTGAGCTT 6957  
DB 6817 AACTATTCAGCATTTTCAAGAGTGGACTCTTTGGAGCCCATGGAAGGTGATGTTGAGCTT 6876  
QY 6958 GAATGGGAAGAAACCACTATGTAATATTTTCAAGCAAAAGATACAAATTCGAAGAGATTT 7017  
DB 6877 GAATGGGAAGAAACCACTATGTAATATTTTCAAGCAAAAGATACAAATTCGAAGAGATTT 6936  
QY 7018 TAAATCCAGGGGCCAAAGTTTACCCCTCATTTCTCCGAATTTGAATTTGCAACCTTAA 7077  
DB 6937 TAAATCCAGGGGCCAAAGTTTACCCCTCATTTCTCCGAATTTGAATTTGCAACCTTAA 6996  
QY 7078 GAAATATCTATGCTTCTCTCACTGTCCTTT 7108  
DB 6997 GAAATATCTATGCTTCTCTCACTGTCCTTT 7027

## RESULT 4

AX686453  
LOCUS  
DEFINITION  
Sequence 9 from Patent WO02057450.  
AX686453  
ACCESSION  
AX686453.1 GI:29372157  
VERSION  
KEYWORDS  
Homo sapiens (human)  
SOURCE  
ORGANISM  
Homo sapiens



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

Edinger, S., Macdougall, J.R., Millet, I., Ellerman, K., Stone, D.J.,  
Gerlach, V., Grosse, W.M., Alsbrook, J.P., Lepley, D.M., Rieger, D.,  
Burgess, C.E., Casman, S.J., Spytek, K.A., Boldog, F.L., Li, L.,  
Padigar, M., Mishra, V., Patturajan, M., Shenoy, S., Rastelli, L.,  
Tchernev, V.T., Vernet, C.A., Zerhusen, B.D., Malyankar, U.M., Guo, Y.,  
Miller, C.E. and Gangolli, E.A.

Proteins and nucleic acids encoding same

Patent: WO 02057450-A 9 25-JUL-2002;

Curagen Corporation (US)

Location/Qualifiers

1..6903

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

## FEATURES

source

## ORIGIN

Query Match 94.0%; Score 6683.6; DB 6; Length 6903;  
Best Local Similarity 98.9%; Pred. No. 0;  
Matches 6842; Conservative 0; Mismatches 19; Indels 57; Gaps 9;  
106 ATGGATTTTCTATCAATTTTCTTTTACCTTTTATTTGGGACTTCAGAGACACAGGTTGAT 165  
1 ATGGATTTTCTATCAATTTTCTTTTACCTTTTATTTGGGACTTCAGAGACACAGGTTGAT 60  
166 GTTTCGAATGCTGCTCTGTTAGTACGATATTAACCAATCTCTTCAATTTCTACAACA 225  
61 GTTTCGAATGCTGCTCTGTTAGTACGATATTAACCAATCTCTTCAATTTCTACAACA 120  
226 TACACCTCACTGTTACTAGAAAGTACAGCAACCAATGTAAACAAACAGGCGCTCCAGTC 285  
121 TACACCTCACTGTTACTAGAAAGTACAGCAACCAATGTAAACAAACAGGCGCTCCAGTC 180  
286 TTCTAGCGGGGAAAGAGTCGATCTGCTGGGATTTCTTCTGTTTGGAAATACACCACT 345  
181 TTCTAGCGGGGAAAGAGTCGATCTGCTGGGATTTCTTCTGTTTGGAAATACACCACT 240  
346 AATCCAAATGGAAGGATATATCTTACATTTGTCMAATATGAAGAGTTTCTCGTGGATG 405  
241 AATCCAAATGGAAGGATATATCTTACATTTGTCMAATATGAAGAGTTTCTCGTGGATG 300  
406 CAACAGTATATACAGAGTCAGATCAAGCCAGACAGTCTGGAAGTTCTTCTTACTAAT 465  
301 CAACAGTATATACAGAGTCAGATCAAGCCAGACAGTCTGGAAGTTCTTCTTACTAAT 360  
466 CTTAATCCTGGAAACACATATGAATTAAGGTTGCTGTGAACACAGTGTGCGATTGGA 525  
361 CTTAATCCTGGAAACACATATGAATTAAGGTTGCTGTGAACACAGTGTGCGATTGGA 420  
526 GTGTTTATGATCCATTTCTCTCCAAA CTGCGAAAGT---GCTCCAGAAAGTGGTG 582  
421 GTGTTTATGATCCATTTCTCTCCAAA CTGCGAAAGTCCAGTCCAGAAAGTGGTG 480  
583 AATCTCAGCTTGGGCTTACAGCTTACAGCTTACAGCTTACAGCTTACCTCGG 642  
481 AATCTCAGCTTGGGCTTACAGCTTACAGCTTACAGCTTACAGCTTACCTCGG 540  
643 CAACCAATGGCAAAATATACAGCTTCAAGTTAGTGTCAAAATGCTCAAGAGTGGGATA 702  
541 CAACCAATGGCAAAATATACAGCTTCAAGTTAGTGTCAAGATGCTCAAGAGTGGGATA 600  
703 GTAGTGAAGATGTCTCAATCAGAGTAGAGGACATTTTGTGCTGGGAAATGGCAGATGC 762  
601 GTAGTGAAGATGTCTCAATCAGAGTAGAGGACATTTTGTGCTGGGAAATGGCAGATGC 660  
763 AAT---GAGAAATAGTGAATCTTTTATGAGTACAGCCAGCTTCTCCAACTTGGT 819  
661 AATGTAGAAATAGTGAATCTTTTATGAGTACAGCCAGCTTCTCCAACTTGGT 720  
820 AGAGTTACACTCCATCGCTTACCAATTCATCAAGCACTGTGACACAGATGAGATC 879

Db QY 721 AGAGTTACACTCCATCCGCTACCAACATTCATCAAGCAAGCTTGACACAGAAATGAGATC 780  
880 AGCTCTGTGTGGAAGAGCCTATCAGTTTGTAGTGACACACTTGAGACCTTTATCAACA 939  
Db 781 AGCTCTGTGTGGAAGAGCCTATCAGTTTGTAGTGACACACTTGAGACCTTTATCAACA 840  
QY 940 TATCTTTTGAAGTTTCAAGTCTGCTCAACTGAAGCAGGTTATTTGATAGTAGTATGTC 999  
Db 841 TATCTTTTGAAGTTTCAAGTCTGCTCAACTGAAGCAGGTTATTTGATAGTAGTATGTC 900  
QY 1000 AGAACACAGAGATCAGTGTGGAAGCACCACCAAACTGCTGAACAGGCAACATCA 1059  
Db 901 AGAACACAGAGATCAGTGTGGAAGCACCACCAAACTGCTGAACAGGCAACATCA 960  
QY 1060 GGAAGTCTCTTTCAATTTTATGGGACCCACCACTATAGTAAACAGGGAATTTAGTTAT 1119  
Db 961 GGAAGTCTCTTTCAATTTTATGGGACCCACCACTATAGTAAACAGGGAATTTAGTTAT 1020  
QY 1120 AGAGTTGAATATATAGGACCATCAGTGTGGAATTTTGGATAACAGCACAAAAGACCTCAAG 1179  
Db 1021 AGAGTTGAATATATAGGACCATCAGTGTGGAATTTTGGATAACAGCACAAAAGACCTCAAG 1080  
QY 1180 TTTGCAATTCATTAACCTTAACACCATTTTACATTTATGATGATGCTATATTTGGGCTGAAACC 1239  
Db 1081 TTTGCAATTCATTAACCTTAACACCATTTTACATTTATGATGATGCTATATTTGGGCTGAAACC 1140  
QY 1240 AGTGCAGGAGCTGGGCGGCAAGTCAATATTTTCAAGTATTCACCTCCACAGATGTTCCAGGG 1299  
Db 1141 AGTGCAGGAGCTGGGCGGCAAGTCAATATTTTCAAGTATTCACCTCCACAGATGTTCCAGGG 1200  
QY 1300 GCAGTGTGTTGATTTTCAACTTGCAGAGGTAGAAATCCAGCAAGTAAAGAAATTTCTTGGAG 1359  
Db 1201 GCAGTGTGTTGATTTTCAACTTGCAGAGGTAGAAATCCAGCAAGTAAAGAAATTTCTTGGAG 1260  
QY 1360 AAAACAAGCAACCAATTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1419  
Db 1261 AAAACAAGCAACCAATTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1320  
QY 1420 ACAGGAATTAATTTTGGAAATTTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1479  
Db 1321 ACAGGAATTAATTTTGGAAATTTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1377  
QY 1480 ATGGCTCCAGAAATTTGTGAACATAGTAGAGCAATTTGTTAGGATTTATATAGGAGTTTCAACA 1539  
Db 1378 ATGGCTCCAGAAATTTGTGAACATAGTAGAGCAATTTGTTAGGATTTATATAGGAGTTTCAACA 1437  
QY 1540 GAGATGTGCTGTGACCTTCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1599  
Db 1438 GAGATGTGCTGTGACCTTCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1497  
QY 1600 AACTTTTCTGCAAGGAATAGAGTGAAGACCAAGCTTCAAGCTTGTAACTTAACTTAACTTAACT 1659  
Db 1498 AACTTTTCTGCAAGGAATAGAGTGAAGACCAAGCTTCAAGCTTGTAACTTAACTTAACTTAACT 1557  
QY 1660 CAGTATATTTACTGACATTTGCAAGCTTCAAGCTTGTAACTTAACTTAACTTAACTTAACTTAACT 1719  
Db 1558 CAGTATATTTACTGACATTTGCAAGCTTCAAGCTTGTAACTTAACTTAACTTAACTTAACTTAACT 1617  
QY 1720 TTTCACTGAGCAATGATTTAGTGTATCTGCTTTTCAATCATGTTGGGAGAGGACCAACA 1779  
Db 1618 TTTCACTGAGCAATGATTTAGTGTATCTGCTTTTCAATCATGTTGGGAGAGGACCAACA 1677  
QY 1780 GTTCTCAGTGTAGGACAGTCCAGAGTCCAGCTTCAAGCTTCAATTTAACTTAACTTAACTTAACT 1839  
Db 1678 GTTCTCAGTGTAGGACAGTCCAGAGTCCAGCTTCAAGCTTCAATTTAACTTAACTTAACTTAACT 1737  
QY 1840 AATATTAGTCTTCTTCACTTATTTTGTATTTTGTATTTGGGATCTTCCAGAAATATCCCAATGGAAA 1899  
Db 1738 AATATTAGTCTTCTTCACTTATTTTGTATTTTGTATTTGGGATCTTCCAGAAATATCCCAATGGAAA 1797  
QY 1900 ATAACTCAGTATGAGTTTATGCAATTTGGAATTTGGATATGAAACAGAGCATTTCCAGATAACT 1959  
Db 1798 ATAACTCAGTATGAGTTTATGCAATTTGGAATTTGGATATGAAACAGAGCATTTCCAGATAACT 1857

QY 1960 ACCATAGATACACGCTTTCTCATTAAC-----AGGGTTAAAGAAATACACAAATACAAA 2013  
DB 1958 ACCATAGATACACGCTTTCTCATTAACAGGTATAGGGTTAAAGAAATACACAAATACAAA 1917  
QY 2014 ATGAGAGTGGCAGCTCAACCCAGATGGAGAAAGTTCTTTGCTGGAAGAAATGACATC 2073  
DB 1918 ATGAGAGTGGCAGCTCAACCCAGTGGAGAAAGTTCTTTGCTGGAAGAAATGACATC 1977  
QY 2074 TTTGTGAGAACTTTCAGAAAGTGAACCGGAATCATCACTCAAGATGTCGGAAGTAATTGAT 2133  
DB 1978 TTTGTGAGAACTTTCAGAAAGTGAACCGGAATCATCACTCAAGATGTCGGAAGTAATTGAT 2037  
QY 2134 GTTACCCGAGATGAATAAGGTTGAAGTGTACCAACCCGAAAGCCCAATGGGATCATTT 2193  
DB 2038 GTTACCCGAGATGAATAAGGTTGAAGTGTACCAACCCGAAAGCCCAATGGGATCATTT 2097  
QY 2194 ATTGCTTATCAAGTGTCTATATAAAATATAGATACCTTATATATAGAACACATCAACA 2253  
DB 2098 ATTGCTTATCAAGTGTCTATATAAAATATAGATACCTTATATATAGAACACATCAACA 2157  
QY 2254 ACAGACATAATATTAAGGAATCTTAAGACCTCACACCTCTATACATTTCTGTAAGTCT 2313  
DB 2158 ACAGACATAATATTAAGGAATCTTAAGACCTCACACCTCTATACATTTCTGTAAGTCT 2217  
QY 2314 TACACGAGATTTGGTCAAGGCAATCAGGTATCTTTTACTCTCTGTAAGGACTTCGGAG 2373  
DB 2218 TACACGAGATTTGGTCAAGGCAATCAGGTATCTTTTACTCTCTGTAAGGACTTCGGAG 2277  
QY 2374 ACTGTGCTCATAGTGCACAGAAATATCACTTACAAAATATTTCTTCGAGAGATT 2433  
DB 2278 ACTGTGCTCATAGTGCACAGAAATATCACTTACAAAATATTTCTTCGAGAGATT 2337  
QY 2434 GAGCTATCATTTCTTCCCAAGTAGTCCCAATGGAAATCATATAAAATATACAAATTTAT 2493  
DB 2338 GAGCTATCATTTCTTCCCAAGTAGTCCCAATGGAAATCATATAAAATATACAAATTTAT 2397  
QY 2494 CTCAGAGAAAGTAAATGGAATGAGGAAGAACTATAAATAACAACCTCTTTTAAACCCAAAAC 2553  
DB 2398 CTCAGAGAAAGTAAATGGAATGAGGAAGAACTATAAATAACAACCTCTTTTAAACCCAAAAC 2457  
QY 2554 ATTAAAGTACTGAAGAAATATACCAATATATCACTTGAAGTGTCTGCTAGTACACATGAAA 2613  
DB 2458 ATT-----CTGAAGAAATATACCAATATATCACTTGAAGTGTCTGCTAGTACACATCAA 2511  
QY 2614 GGTGAAGGAGTTCGGAGTGTCTCCATAAGTATATCTGACGGAGGAAGATGCTCTGATTTCT 2673  
DB 2512 GGTGAAGGAGTTCGGAGTGTCTCCATAAGTATATCTGACGGAGGAAGATGCTCTGATTTCT 2571  
QY 2674 CCCCCTCAAGACTTCTCTGTAAGCAAGTGTCTGGTGTCAACGGTGAAGTTGTCTATGGCAA 2733  
DB 2572 CCCCCTCAAGACTTCTCTGTAAGCAAGTGTCTGGTGTCAACGGTGAAGTTGTCTATGGCAA 2631  
QY 2734 CCACCCCTGGAGCAAAATGGAATTTCTTTATATACACAGTTTATGCTCT-----GGAATAGA 2790  
DB 2632 CCACCCCTGGAGCAAAATGGAATTTCTTTATATACACAGTTTATGCTCTGGAGGAATAGA 2691  
QY 2791 TCATCATTTAAACTATTATATGTCACCTGAAAATCATCATTTGGAGTTATCAGATTTGGATTAT 2850  
DB 2692 TCATCATTTAAACTATTATATGTCACCTGAAAATCATCATTTGGAGTTATCAGATTTGGATTAT 2751  
QY 2851 AATGTTGAAATACAGTGTCTTATGTAACAGCTAGCACCAAGATTTGGTGTATGGGAAAAACAGGA 2910  
DB 2752 AATGTTGAAATACAGTGTCTTATGTAACAGCTAGCACCAAGATTTGGTGTATGGGAAAAACAGGA 2811  
QY 2911 AGCAATATCATTTAGCTTTTAAACACAGAGGGAGCAACGAGATCCTCCCAAGATGTT 2970  
DB 2812 AGCAATATCATTTAGCTTTTAAACACAGAGGG-----ACCAAGCGATCCTCCCAAGATGTT 2868  
QY 2971 TATTATGCAAACTCAGTGTCTTCAATCAATAATTTCTTTCTGGACACCTCTCTCAAAACCT 3030  
DB 2869 TATTATGCAAACTCAGTGTCTTCAATCAATAATTTCTTTCTGGACACCTCTCTCAAAACCT 2928

QY 3031 AATGGGATTATACAAATATTACTCTCTGTTTATTAACAGAAATACCTTCAGGTACTTTTATGCGAG 3090  
DB 2929 AATGGGATTATACAAATATTACTCTCTGTTTATTAACAGAAATACCTTCAGGTACTTTTATGCGAG 2988  
QY 3091 AATTTTACATCTCATGAACTTAACCAATGACTTTTGACAAATATGACTCTATCCACAAATTATA 3150  
DB 2989 AATTTTACATCTCATGAACTTAACCAATGACTTTTGACAAATATGACTCTATCCAAATTTATA 3048  
QY 3151 GATAAACTGCAATATTTCAGCTACTATATACATTTTGGTTTAAACAGCAAGTACTTTCAGTTGGA 3210  
DB 3049 GATAAACTGCAATATTTCAGCTACTATATACATTTTGGTTTAAACAGCAAGTACTTTCAGTTGGA 3108  
QY 3211 AATGGGAATAAAGCAGTGCATCATTTGAAGTATATACAGATCAAGACATACCTGAAGGG 3270  
DB 3109 AATGGGAATAAAGCAGTGCATCATTTGAAGTATATACAGATCAAGACGCTCCCTGAAGGG 3168  
QY 3271 TTTTGTGGAACCTGACTTTACGAATCCATTTCTGCTCAACTGCAATTAATGTAAGCTGGGTC 3330  
DB 3169 TTTTGTGGAACCTGACTTTACGAATCCATTTCTGCTCAACTGCAATTAATGTAAGCTGGGTC 3328  
QY 3331 CCACCGGCTCAACCAACCGGTCTAGTCTTCTACTATGTTTCTACTGATCTTTACAGCAGACT 3390  
DB 3229 CCACCGGCTCAACCAACCGGTCTAGTCTTCTACTATGTTTCTACTGATCTTTACAGCAGACT 3288  
QY 3391 CCTCGCATGTGAGACCACTCTTGTGTAATATGAGAGAGGACATATATTTGATATATCTG 3450  
DB 3289 CCTCGCATGTGAGACCACTCTTGTGTAATATGAGAGAGGACATATATTTGATATATCTG 3348  
QY 3451 GAAAATACACCTGATTATATATTTAAATTTACTCCATCAACAGAAAGGATTTCTCTGAT 3510  
DB 3349 GAAAATACACCTGATTATATATTTAAATTTACTCCATCAACAGAAAGGATTTCTCTGAT 3408  
QY 3511 ACCTATPACTGCCAGCTATACATCAAGACTGAAGAAGATGTCTCCAGAAAATTTCCACCAATA 3570  
DB 3409 ACCTATPACTGCCAGCTATACATCAAGACTGAAGAAGATGTCTCCAGAAAATTTCCACCAATA 3468  
QY 3571 ATCAACACTTTTTTAAACCTTTCTCTACTCTGATTTCTTTATCATGGGATCCCCAGTA 3630  
DB 3469 ATCAACACTTTTTTAAACCTTTCTCTACTCTGATTTCTTTATCATGGGATCCCCAGTA 3528  
QY 3631 AAGCCAAATGGTGCAATATATAGTTATGATTTAACTTTTCAAGGACCAATGAAATTTAT 3690  
DB 3529 AAGCCAAATGGTGCAATATATAGTTATGATTTAACTTTTCAAGGACCAATGAAATTTAT 3588  
QY 3691 TCTTTTCACTTCTTGATAATTTACATAATATTTGGAAGAGCTTTTACCAATTTTACATTTAT 3750  
DB 3589 TCTTTTCACTTCTTGATAATTTACATAATATTTGGAAGAGCTTTTACCAATTTTACATTTAT 3648  
QY 3751 AGCTTTTTTGTGCGCAAGAACTAGAAAAGGACTTTGGTCTCTTCCAGATTTCTTTTCTTT 3810  
DB 3649 AGCTTTTTTGTGCGCAAGAACTAGAAAAGGACTTTGGTCTCTTCCAGATTTCTTTTCTTT 3708  
QY 3811 TACACAGATGAGTGCAGTCCGTTAGCACTTCCACAAAATTTTACATTTTACACTGTACT 3870  
DB 3709 TACACAGATGAGTGCAGTCCGTTAGCACTTCCACAAAATTTTACATTTTACACTGTACT 3768  
QY 3871 TCACAGCTTTGTATGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGGTATTTGTTAAAGTA 3930  
DB 3769 TCACAGCTTTGTATGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGGTATTTGTTAAAGTA 3828  
QY 3931 TATAGTTTTTAAATTCATGAACATGAACTGACATATATATTTAAGAAATATATACAGGA 3990  
DB 3829 TATAGTTTTTAAATTCATGAACATGAACTGACATATATTTAAGAAATATATACAGGA 3888  
QY 3991 TTTTAAACTGGAACCAACTTTGTTGCACTGGAAACAGTCAGCACCTACTCTATCCGTGTA 4050  
DB 3889 TTTTAAACTGGAACCAACTTTGTTGCACTGGAAACAGTCAGCACCTACTCTATCCGTGTA 3948  
QY 4051 TCTGCTTTTCAACCAAGTTGGAAATGCGCAATCAATTTAGTAAATGTAAGTAAATTTCAACAC 4110  
DB 3949 TCTGCTTTTCAACCAAGTTGGAAATGCGCAATCAATTTAGTAAATGTAAGTAAATTTCAACAC 4008  
QY 4111 CAAGAATCAGTTTCCAGATGTCTGTCAGAAATATATGCAAGTGCATGGCAACCTAGCTGGCAGTCA 4170

[illegible]





Db 3085 ACAGCAAGTACTCTCAGTTGGAAATGGGNAATAAAGCAGTGACATCTTGAAGTATACACA 3144  
Qy 3250 GATCAAGACATACCTGGAAGGTTTGGTGGAAACCTGACTTACGAATCAATTCGTCAACT 3309  
Db 3145 GATCAAGACGATACCTGGAAGGTTTGGTGGAAACCTGACTTACGAATCAATTCGTCAACT 3204  
Qy 3310 GCAATAAATGTAAGCTGGTCCACCGCTCAACCAACGGCTAGTCTCTACTATGTT 3369  
Db 3205 GCAATAAATGTAAGCTGGTCCACCGCTCAACCAACGGCTAGTCTCTACTATGTT 3264  
Qy 3370 TCACTGATCTTACAGCAGACTCTCTGCACTGTGAGACACCTCTTGTATCATATGAGAGA 3429  
Db 3265 TCACTGATCTTACAGCAGACTCTCTGCACTGTGAGACACCTCTTGTATCATATGAGAGA 3324  
Qy 3430 AGCATATATTTGATATCTGGAAAAATACATGATATATATAATAAATAATCTCCATCA 3489  
Db 3325 AGCATATATTTGATATCTGGAAAAATACATGATATATATAATAAATAATCTCCATCA 3384  
Qy 3490 ACAGAAAGGGATCTCTGTATACCTATACCTGCGCAGCTATACATCAAGACTGAGAGAT 3549  
Db 3385 ACAGAAAGGGATCTCTGTATACCTATACCTGCGCAGCTATACATCAAGACTGAGAGAT 3444  
Qy 3550 GTCCACAGAACTTCAACCAATAATCAACACTTTTAAAAACCTTCTCTACTCAGTTCTC 3609  
Db 3445 ATCCACAGAACTTCAACCAATAATCAACACTTTTAAAAACCTTCTCTACTCAGTTCTC 3504  
Qy 3610 TTATATGGGATCCCCAGTAAAGCCAAATGGTGCATAATAAGTTATGATTAATTAACTTTA 3669  
Db 3505 TTATATGGGATCCCCAGTAAAGCCAAATGGTGCATAATAAGTTATGATTAATTAACTTTA 3564  
Qy 3670 CAAGGACCAATGAAATTAATCTTTCATTTACTTCTGATATTAATTAATTAATTAAGTTA 3729  
Db 3565 CAAGGACCAATGAAATTAATCTTTCATTTACTTCTGATTAATTAATTAATTAAGTTA 3624  
Qy 3730 CTTTCCACATTTACATTAATATAGCTTTTGTGCGCCAGAACTAGAAAGGACTTGGT 3789  
Db 3625 CTTTCCACATTTACATTAATATAGCTTTTGTGCGCCAGAACTAGAAAGGACTTGGT 3684  
Qy 3790 CTTTCCAGTATCTTTTCAACAGATGAGTCAAGTGCCTGTTAGCACCTCCACAAAT 3849  
Db 3685 CTTTCCAGTATCTTTTCAACAGATGAGTCAAGTGCCTGTTAGCACCTCCACAAAT 3744  
Qy 3850 TTGACTTTAATCAACTGATCTCAGACTTTGTATGGCTGAATGGAGCCCAAGTCTCTTT 3909  
Db 3745 TTGACTTTAATCAACTGATCTCAGACTTTGTATGGCTGAATGGAGCCCAAGTCTCTTT 3804  
Qy 3910 CCAGGTGGTATGTTAAAGTATATAGTTTAAATTCATGAACATGAACCTGACACTATA 3969  
Db 3805 CCAGGTGGTATGTTAAAGTATATAGTTTAAATTCATGAACATGAACCTGACACTATA 3864  
Qy 3970 TATTATAAGAAATATATCAGGATTTAAACTGAAGCCAAACTTGTGGACTGGAAACGATC 4029  
Db 3865 TATTATAAGAAATATATCAGGATTTAAACTGAAGCCAAACTTGTGGACTGGAAACGATC 3924  
Qy 4030 AGCACTTACTCTATCCGTGTATCTGGTTCACCAAGTTGGAAATGGCAATCAATTTAGT 4089  
Db 3925 AGCACTTACTCTATCCGTGTATCTGGTTCACCAAGTTGGAAATGGCAATCAATTTAGT 3984  
Qy 4090 AATGTAGTAAATTCACAAACCAAGAAATCAGTTTCCAGATGCTGTCGAGATATGCAAGTGC 4149  
Db 3985 AATGTAGTAAATTCACAAACCAAGAAATCAGTTTCCAGATGCTGTCGAGATATGCAAGTGC 4044  
Qy 4150 ATGGCAACTAGCTGGCAGTCAAGTTTGTGAAATGGGATCCACCCAAAGGCAAAATGGA 4209  
Db 4045 ATGGCAACTAGCTGGCAGTCAAGTTTGTGAAATGGGATCCACCCAAAGGCAAAATGGA 4104  
Qy 4210 ATAATAACCGAGTATATGTTAAAGTAAAGGAAATTTCTACAAAGTTTCTCCCAAGAT 4269  
Db 4105 ATAATAACCGAGTATATGTTAAAGTAAAGGAAATTTCTACAAAGTTTCTCCCAAGAT 4164  
Qy 4270 CACATGTACTTTTATTAAGCTTTCTGCAATACCTCATATGCTTTTAAAGTAAAGCT 4329  
Db 4165 CACATGTACTTTTATTAAGCTTTCTGCAATACCTCATATGCTTTTAAAGTAAAGCT 4224

Qy 4330 TCAACCTCAGCTGGTGAAGGTGATGAAGACATGCAATGTCAGCACTACTCCTGAAACA 4389  
Db 4225 TCAACCTCAGCTGGTGAAGGTGATGAAGACATGCAATGTCAGCACTACTCCTGAAACA 4284  
Qy 4390 GTTCCCAAGTGTCCCAACATATTTGCTTTCTGATGTTTCAGTCAACTAGTGAACATTTG 4449  
Db 4285 GTTCCCAAGTGTCCCAACATATTTGCTTTCTGATGTTTCAGTCAACTAGTGAACATTTG 4344  
Qy 4450 ACATGGATAAGACTCTGACACTATCTTTGGCTACTTTCAAAATTAACAAATTAACACTCAA 4509  
Db 4345 ACATGGATAAGACTCTGACACTATCTTTGGCTACTTTCAAAATTAACAAATTAACACTCAA 4404  
Qy 4510 CTTCTGCTCAAAATTCGAAAGATCGGAATCGGAATCGGAATCGGTGGAATATCAAAATTT 4569  
Db 4405 CTTCTGCTCAAAATTCGAAAGATCGGAATCGGAATCGGTGGAATATCAAAATTT 4464  
Qy 4570 CAATACCTCTATGAAGCTCACTTAACTGAAGACAGATATATGATTAAGAAATTTAGA 4629  
Db 4465 CAATACCTCTATGAAGCTCACTTAACTGAAGACAGATATATGATTAAGAAATTTAGA 4524  
Qy 4630 TGGTATAGATTTCAAGTGGCTGCGACCACTGCTGGCTATGCGCAATGCTTCAAACTGG 4689  
Db 4525 TGGTATAGATTTCAAGTGGCTGCGACCACTGCTGGCTATGCGCAATGCTTCAAACTGG 4584  
Qy 4690 ATTTCTCAAAATCTGCTGCTGCTCCAGATGGTCTCTCGAATGTTTCATGATGA 4749  
Db 4585 ATTTCTCAAAATCTGCTGCTGCTCCAGATGGTCTCTCGAATGTTTCATGATGA 4644  
Qy 4750 GCAACATCACTTTTAGCATCAGCATAGCTGGAGTGAACCTGCTCATTAAGTGAACCA 4809  
Db 4645 GCAACATCACTTTTAGCATCAGCATAGCTGGAGTGAACCTGCTCATTAAGTGAACCA 4704  
Qy 4810 ACATGTTATCTGATGATGTCGCAATCGGTAGATTAATGATGAATTTAATATATCTTCAATC 4869  
Db 4705 ACATGTTATCTGATGATGTCGCAATCGGTAGATTAATGATGAATTTAATATATCTTCAATC 4764  
Qy 4870 AAGTCAAAATGAAGAAATAAAAACCATAGAAATTTAAAGATTTAGAAATTTTCAAGATAT 4929  
Db 4765 AAGTCAAAATGAAGAAATAAAAACCATAGAAATTTAAAGATTTTAGAAATTTTCAAGATAT 4824  
Qy 4930 TCTGTAGTATCACTGCTATTTACTGGAACATTTAGTGTGTCATATGATAGAGGAAAGTCA 4989  
Db 4825 TCTGTAGTATCACTGCTATTTACTGGAACATTTAGTGTGTCATATGATAGAGGAAAGTCA 4884  
Qy 4990 AGTGTGGAATGATTTACTTACTTCTAGAAATCAGCCCAAGAGCCACCTAACCAACATG 5049  
Db 4885 AGTGTGGAATGATTTACTTACTTCTAGAAATCAGCCCAAGAGCCACCTAACCAACATG 4944  
Qy 5050 ACATTTCAGAGATACAGATGAAGTTACAAATTTTCAATTTAACTTCTCTCTCTCTCT 5109  
Db 4945 ACATTTCAGAGATACAGATGAAGTTACAAATTTTCAATTTAACTTCTCTCTCTCTCT 5004  
Qy 5110 CAACCTTAATGGAATATCAAGTATATCAAGTCTGGTTTACCGAGAAGATGATCTACT 5169  
Db 5005 CAACCTTAATGGAATATCAAGTATATCAAGTCTGGTTTACCGAGAAGATGATCTACT 5064  
Qy 5170 GCTGTCCAGATTCACAACTCAGTATTTATACAGAAACCAACACATTTCTGTTGCAATG 5229  
Db 5065 GCTGTCCAGATTCACAACTCAGTATTTATACAGAAACCAACACATTTCTGTTGCAATG 5124  
Qy 5230 CTAGAAGGACTAAAAGGTGGACATACATACATATCAGTGTGTTTACGAGTCAATAGTGT 5289  
Db 5125 CTAGAAGGACTAAAAGGTGGACATACATACATATCAGTGTGTTTACGAGTCAATAGTGT 5184  
Qy 5290 GGTGAGGTTCCAAAGGTTCCGATGAGATAAACATGAGATATCAAGCTCAGGACGACCA 5349  
Db 5185 GGTGAGGTTCCAAAGGTTCCGATGAGATAAACATGAGATATCAAGCTCAGGACGACCA 5244  
Qy 5350 AAAACCAACCAACCCCTTATTTATGATGTCACAGGAAACTGCTGTGATCTTCAACAA 5409  
Db 5245 AAAACCAACCAACCCCTTATTTATGATGTCACAGGAAACTGCTGTGATCTTCAACAA 5304



|    |      |   |      |                          |   |   |                            |
|----|------|---|------|--------------------------|---|---|----------------------------|
| QY | 5410 | ATTACAAATCAGAAATGCAATATGTTACTACAGTGATGATCATGGACCAATAAAAAATGTA | 5469 | Db                       | 6382  | GATGTTCAATAGATTGGACTATCAGGGATCTGAAATTTGAAGGCATGGGATTCGATG       | 6441                       |
| Db | 5305 | ATTACAAATCAGAAATGCAATATGTTACTACAGTGATGATCATGGACCAATAAAAAATGTA | 5364 | QY                       | 6550  | ACTGTTTCGACAGTGTAACCTTTACTCCCTGGCCAGAGCATGGGTTCTCTGAGAAACAGCGCC | 6609                       |
| QY | 5470 | CAAGTCTCTCGACAGAAACAGAGGCTCAGCATGATGGAATGTAAACAAAGTGGTATGAT   | 5529 | Db                       | 6442  | ACTGTTTCGACAGTGTAACCTTTACTCCCTGGCCAGAGCATGGGTTCTCTGAGAAACAGCGCC | 6501                       |
| Db | 5365 | CAAGTCTCTCGACAGAAACAGAGGCTCAGCATGATGGAATGTAAACAAAGTGGTATGAT   | 5424 | QY                       | 6610  | CCTCTAAATTCACCTTTGTGAAGTTGGTTCGAGCAAGCAGGGCACAATGACACCACTATG    | 6669                       |
| QY | 5530 | GCATATTTTAAATGAACAGAGCCATATTTTACAAATGGAAGCTTTCCTAACCCCTCATCT  | 5589 | Db                       | 6502  | CCTCTAAATTCACCTTTGTGAAGTTGGTTCGAGCAAGCAGGGCACAATGACACCACTATG    | 6561                       |
| Db | 5425 | GCATATTTTAAATGAACAGAGCCATATTTTACAAATGGAAGCTTTCCTAACCCCTCATCT  | 5484 | QY                       | 6670  | ATTGTTCTACAGTGCTGGAGTTGGAAGAACTGAGATTTTATTGCTCTGACCACTATTA      | 6729                       |
| QY | 5590 | ACAGAAGGAAGACAAAGTTTGTGCAATGAAGAAATCTACATCATAGTGCTGATAT       | 5649 | Db                       | 6562  | ATTGTTCTACAGTGCTGGAGTTGGAAGAACTGAGATTTTATTGCTCTGACCACTATTA      | 6621                       |
| Db | 5485 | ACAGAAGGAAGACAAAGTTTGTGCAATGAAGAAATCTACATCATAGTGCTGATAT       | 5544 | QY                       | 6730  | ACACAACATATAAATGACCATGATTTTGTGGATATATATGACTAGTAGTGAATGAGA       | 6789                       |
| QY | 5650 | GCATGATGATTCCTCGCAATGAAGACAAATTTGCAATGGACCACTGAAACCAAAAAAG    | 5709 | Db                       | 6622  | ACACAACATATAAATGACCATGATTTTGTGGATATATATGACTAGTAGTGAATGAGA       | 6681                       |
| Db | 5545 | GCATGATGATTCCTCGCAATGAAGACAAATTTGCAATGGACCACTGAAACCAAAAAAG    | 5604 | QY                       | 6790  | AGTGAAGAAATGTCATGCTGAGCAATCTGGCAGAGTATATCTTTTTTACACCACTGAT      | 6849                       |
| QY | 5710 | CAATCTATTTAAATTTAGAGCTACAAATATTTATGGGCAATTTTACTGACTGATTA      | 5769 | Db                       | 6682  | AGTGAAGAAATGTCATGCTGAGCAATCTGGCAGAGTATATCTTTTTTACACCACTGAT      | 6741                       |
| Db | 5605 | CAATCTATTTAAATTTAGAGCTACAAATATTTATGGGCAATTTTACTGACTGATTA      | 5664 | QY                       | 6850  | CTGGATCTCTTATCAAAATGAAGAAAGTAAATCAGCCCATCTGTTTGTAACTATTTCAGCA   | 6909                       |
| QY | 5770 | TCTGACCTGTTAAGACTTTTAGGGGAAGACTTTTCAAGAAAGACCCGTAGAGATCATCTT  | 5829 | Db                       | 6742  | CTGGATCTCTTATCAAAATGAAGAAAGTAAATCAGCCCATCTGTTTGTAACTATTTCAGCA   | 6801                       |
| Db | 5665 | TCTGACCTGTTAAGACTTTTAGGGGAAGACTTTTCAAGAAAGACCCGTAGAGATCATCTT  | 5724 | QY                       | 6910  | CTTCAGAGATGAGCTCTTTGGACGCCATGGAAAGTAAATCTGATGTTGAGCTTCAATGGGAA  | 6966                       |
| QY | 5830 | TCCGTCACTTTGTGATCCTTTCAATTAATCTCCTTTGGAAACAGCTATTTTGGATTGCA   | 5889 | Db                       | 6802  | CTTCAGAGATGAGCTCTTTGGACGCCATGGAAAGTAAATCTGATGTTGAGCTTCAATGGGAA  | 6861                       |
| Db | 5725 | TCCGTCACTTTGTGATCCTTTCAATTAATCTCCTTTGGAAACAGCTATTTTGGATTGCA   | 5784 | QY                       | 6967  | GAACCACTATGTAATATTCAGACCAAGAGTACAATTTGGAAGAGATTTTAAATCCCA       | 7026                       |
| QY | 5890 | AGAAATTCGACAGAGCAGAAAGAGCTGGACATCTCTCTCAGAGTCAGAAATTA         | 5949 | Db                       | 6862  | GAACCACTATGTAATATTCAGACCAAGAGTACAATTTGGAAGAGATTTTAAATCCCA       | 6921                       |
| Db | 5785 | AGAAATTCGACAGAGCAGAAAGAGCTGGACATCTCTCTCAGAGTCAGAAATTA         | 5844 | QY                       | 7027  | GGGGCCAAAGTTACCCCTCATCTTCGAAATGAAATGCAACCTTAAAGAAATATCT         | 7086                       |
| QY | 5950 | GACACTAAATTTGAAGTGGATCAGCTCATCAGAGTGGCAGACCTCGAACTGAAGACGAG   | 6009 | Db                       | 6922  | GGGGCCAAAGTTACCCCTCATCTTCGAAATGAAATGCAACCTTAAAGAAATATCT         | 6981                       |
| Db | 5845 | GACACTAAATTTGAAGTGGATCAGCTCATCAGAGTGGCAGACCTCGAACTGAAGACGAG   | 5904 | RESULT 6                 |   |   |                            |
| QY | 6010 | AGATTAACGGCCCAATGAAGCAAGAAATCTCTCGCAACATGTTGAAGAGCTTTGCACA    | 6069 | LOCUS                    | AX751415  | 5877 bp   | DNA linear PAT 20-JUN-2003 |
| Db | 5905 | AGATTAACGGCCCAATGAAGCAAGAAATCTCTCGCAACATGTTGAAGAGCTTTGCACA    | 5964 | DEFINITION               | Sequence 7 from Patent WO03033688.                |   |                            |
| QY | 6070 | AACAACTAAATTTGAAGAGAAATTTTCCGAATTTACCAAAATTTCTTCAGGATCTT      | 6129 | ACCESSION                | AX751415  |   |                            |
| Db | 5965 | AACAACTAAATTTGAAGAGAAATTTTCCGAATTTACCAAAATTTCTTCAGGATCTT      | 6024 | VERSION                  | AX751415.1  | GI:32133734   |                            |
| QY | 6130 | TCTTCAACTGATGCTGATCTGCTTGAATGAGCAAAACCGTTTCCCAACATAAAA        | 6189 | KEYWORDS                 | Homo sapiens (human)                              |   |                            |
| Db | 6025 | TCTTCAACTGATGCTGATCTGCTTGAATGAGCAAAACCGTTTCCCAACATAAAA        | 6084 | SOURCE                   | Homo sapiens                                      |   |                            |
| QY | 6190 | CCATATAATTAATAACAGAGTAAGCTGATGAGCTGACGCTAGTGTTCAGGTTCCGAT     | 6249 | ORGANISM                 | Homo sapiens                                      |   |                            |
| Db | 6085 | CCAT---ATAATAACAGAGTAAGCTGATGAGCTGACGCTAGTGTTCAGGTTCCGAT      | 6141 | REFERENCE                | 1   |   |                            |
| QY | 6250 | TATATTAATCCAGCTATATTTCTGGTTATTTATGTCCTCAATGAATTTATGCTCTCAA    | 6309 | AUTHORS                  | Liou, J.R.  |   |                            |
| Db | 6142 | TATATTAATCCAGCTATATTTCTGGTTATTTATGTCCTCAATGAATTTATGCTCTCAA    | 6201 | TITLE                    | Regulation of human receptor tyrosine phosphatase |   |                            |
| QY | 6310 | GGTCCACTACAGGAACAGTTGGAGATTTTGGAGAAATGTTGGGAAACACGGGCAAAA     | 6369 | JOURNAL                  | Patent: WO 03033688-A 7 24-APR-2003;              |   |                            |
| Db | 6202 | GGTCCACTACAGGAACAGTTGGAGATTTTGGAGAAATGTTGGGAAACACGGGCAAAA     | 6261 | FEATURES                 | Bayer Aktiengesellschaft (DE)                     |   |                            |
| QY | 6370 | ACATTAGTAATGCTAAACACAGTGTTTTGAAGAAAGGACGGATCAGATGCCATCAGTATGG | 6429 | source                   | 1..5877   |   |                            |
| Db | 6262 | ACATTAGTAATGCTAAACACAGTGTTTTGAAGAAAGGACGGATCAGATGCCATCAGTATGG | 6321 | /organism="Homo sapiens" | /mol_type="unassigned DNA"                        |   |                            |
| QY | 6430 | CCAGAGACAAACAGCCAGTACTGCTTTGGAGATATAGTATTAACAAAGCTAATGGAG     | 6489 | /db_xref="taxon:9606"    |   |   |                            |
| Db | 6322 | CCAGAGACAAACAGCCAGTACTGCTTTGGAGATATAGTATTAACAAAGCTAATGGAG     | 6381 | ORIGIN                   |   |   |                            |
| QY | 6490 | GATGTTCAAATAGATTGCACTATCAGGGATCTCGAAATTTGAAGGCGATGGGATTCGATG  | 6549 | Query Match              | 79.1%;  | Score 5619.4;   | DB 6; Length 5877;         |
|    |      |   |      | Best Local Similarity    | 97.6%;  | Pred. No. 0;  |                            |
|    |      |   |      | Matches 5736;            | Conservative 0;                                   | Mismatches 111;   | Indels 30; Gaps 2;         |
|    |      |   |      | QY                       | 1129  | TTATATGGACCATCAGGTGCGATTTTGGATAACAGCAAAAGACCTCAAGTTTCATTC       | 1188                       |



Db 4 TTAGGAGGAAACTAACAACAGAAAGGACATCCACCAAAAACCCATCTGTACATCAC 63  
QY 1189 ACTAACCTAACCACTTTACAATGTATGATGTCTATATTGGCGCTGAAACCAAGTGCAGGG 1248  
Db 64 CATCATCAAGACCAAAAGTAGATAAAAACCAAAAGATGGGAAAAACAGAGCAAAA 123  
QY 1249 ACTGGCCCCAAGTCAATATTTCAGTATTTCATCTCCACAGATGTTCCAGGCGAGTGT 1308  
Db 124 ACTGGAAACTCTAAAAGCAGAGCACCTCTCTCTCCCAAGGATGGCAGGGCAGTGT 183  
QY 1309 GATTTACAACTTGCAGAGGTAGAAATCCAGCAAGTAAAGAAATACCTTGGAGAAACCAACGA 1368  
Db 184 GATTTACAACTTGCAGAGGTAGAAATCCAGCAAGTAAAGAAATACCTTGGAGAAACCAACGA 243  
QY 1369 CAACCAAAATGGAATTAATAACCAATACCGAGTAAAGTCTAGTTCACAGACAGGAATA 1428  
Db 244 CAACCAAAATGGAATTAATAACCAATACCGAGTAAAGTCTAGTTCACAGACAGGAATA 303  
QY 1429 ATTTTGGAAATACCTTGTCTCACTGGAATTAATGAGTATATAAATGACCCCATGGCTCCA 1488  
Db 304 ATTTTGGAAATACCTTGTCTCACTGGAATTAATGAGTATATAAATGACCCCATGGCTCCA 363  
QY 1489 GAAATGTCAACATAGTAGCAATGAGTATATATGAGGTTTACAGACAGATGTG 1548  
Db 364 GAAATGTCAACATAGTAGCAATGAGTATATATGAGGTTTACAGACAGATGTG 423  
QY 1549 TCTGACCTTCACTCACTTCTCACTTATATATAACAGCCATCCAGATAAACTTCT 1608  
Db 424 TCTGACCTTCACTCACTTCTCACTTATATAACAGCCATCCAGATAAACTTCT 483  
QY 1609 GAAGGATAGAGCTGAAGACCGAGCTTCAACGTTGTAACTACAGGAATCAGTATATT 1668  
Db 484 GAAGGATAGAGCTGAAGACCGAGCTTCAACGTTGTAACTACAGGAATCAGTATATT 543  
QY 1669 ACTGACATTCAGCTGAACAGCTGTCTATGTTATCAGGAGACTTGTACTTTCACTGAG 1728  
Db 544 ACTGACATTCAGCTGAACAGCTGTCTATGTTATCAGGAGACTTGTACTTTCACTGAG 603  
QY 1729 CACATGATTAGTATCTGCTTTTCACTATGAGGAGAGGACCAACAGTTCTCAGT 1788  
Db 604 CACATGATTAGTATCTGCTTTTCACTATGAGGAGAGGACCAACAGTTCTCAGT 663  
QY 1789 GTTAGGACAGCTCAGCAAGTGCAGCTCCATTAAATTTATAAATTAATAAATATTAGT 1848  
Db 664 GTTAGGACAGCTCAGCAAGTGCAGCTCCATTAAATTTATAAATTAATAAATATTAGT 723  
QY 1849 TCTTCACTATTTTGTATTATTTGGATCTCTCAGATATCCCAATGGAAATTAATCTCAC 1908  
Db 724 TCTTCACTATTTTGTATTATTTGGATCTCTCAGATATCCCAATGGAAATTAATCTCAC 783  
QY 1909 TATACGATTTATGCAATGGAAATGGATACAAACAGAGCAATCCAGATACTACCATAGAT 1968  
Db 784 TATACGATTTATGCAATGGAAATGGATACAAACAGAGCAATCCAGATACTACCATAGAT 843  
QY 1969 AACAGTTTCTCATACAGGTTAAAGAAATACAAAATACAAAATGAGAGTGGCAGCC 2028  
Db 844 AACAGTTTCTCATACAGGTTAAAGAAATACAAAATACAAAATGAGAGTGGCAGCC 903  
QY 2029 TCAACCCAGATGGAGAAATTTCTTGTCTGAAAGAAATGACATCTTTGTGAGAACTTCA 2088  
Db 904 TCAACCCAGATGGAGAAATTTCTTGTCTGAAAGAAATGACATCTTTGTGAGAACTTCA 963  
QY 2089 GAAGATGAACCCGAATCATCACTCAAGATGTGGAATTAATGATGTTTACCCGAGATGAA 2148  
Db 964 GAAGATGAACCCGAATCATCACTCAAGATGTGGAATTAATGATGTTTACCCGAGATGAA 1023  
QY 2149 ATAAGTTGAAGTGTCAACCCCGAAAGCCCAATGGGATCATTTATGCTTTATGAAGTG 2208  
Db 1024 ATAAGTTGAAGTGTCAACCCCGAAAGCCCAATGGGATCATTTATGCTTTATGAAGTG 1083  
QY 2209 CTATATAAATAATATAGATCTTTTATATATGAAGAACACATCAACAGACATATATTA 2268  
Db 1084 CTATATAAATAATATAGATCTTTTATATATGAAGAACACATCAACAGACATATATTA 1143

QY 2269 AGGAACTTAAGACCTTCACACCTCTATATAACATTTCTGTAAAGTCTTTACACCAAGTTTGGT 2328  
Db 1144 AGGAACTTAAGACCTTCACACCTCTATATAACATTTCTGTAAAGTCTTTACACCAAGTTTGGT 1203  
QY 2329 CATGGCAATCAGGTATCTTCTTACTCTCTGTAAAGACTTCGGAGACTTGGCTGTATAGT 2388  
Db 1204 CATGGCAATCAGGTATCTTCTTACTCTCTGTAAAGACTTCGGAGACTTGGCTGTATAGT 1263  
QY 2389 GCACAGAAAAATATCACTTACAAAAATATTTCTTCTGGAGAGATTGAGCTATCATTTCTTT 2448  
Db 1264 GCACAGAAAAATATCACTTACAAAAATATTTCTTCTGGAGAGATTGAGCTATCATTTCTTT 1323  
QY 2449 CCCCAGAGTAGTCCCAATGGAAATCATAAAAATATACAATTTATCTCAAGAGAAATAT 2508  
Db 1324 CCCCAGAGTAGTCCCAATGGAAATCATAAAAATATACAATTTATCTCAAGAGAAATAT 1383  
QY 2509 GGAATATGAGGAAGAACTATTAATCAACCTCTTTAAACCCAAAAACATTAAGTACTGAAG 2568  
Db 1384 GGAATATGAGGAAGAACTATTAATCAACCTCTTTAAACCCAAAAACATTAAGTACTGAAG 1443  
QY 2569 AAATATACCAATATATCACTGAGGTGTCTGTAGTACACTCAAAAGGTGAAGGTTCCG 2628  
Db 1444 AAATATACCAATATATCACTGAGGTGTCTGTAGTACACTCAAAAGGTGAAGGTTCCG 1503  
QY 2629 AGTGTCTCCCATTAAGTATATCTGACGGAAGATGTCTCTGTATTTCTCCCTCAAGACTTC 2688  
Db 1504 AGTGTCTCCCATTAAGTATATCTGACGGAAGATGTCTCTGTATTTCTCCCTCAAGACTTC 1563  
QY 2689 TCTGTAAAAACAGTTGTCTGGTGTACGGTGAAGTTGTCTATGCAACCAACCCCTGGAGCCA 2748  
Db 1564 TCTGTAAAAACAGTTGTCTGGTGTACGGTGAAGTTGTCTATGCAACCAACCCCTGGAGCCA 1623  
QY 2749 AATGCAATTTATCTTTTATACACAGTTTATGTCTGGAATAGATCATTAATAAACTATT 2808  
Db 1624 AATGCAATTTATCTTTTATACACAGTTTATGTCTGGAATAGATCATTAATAAACTATT 1583  
QY 2809 AATGTCACTGAAACATCAATGGAGTTATCAGATTTTGGATTAATAATGTTGAATACAGTGT 2868  
Db 1684 AATGTCACTGAAACATCAATGGAGTTATCAGATTTTGGATTAATAATGTTGAATACAGTGT 1743  
QY 2869 TATGTAAACAGCTAGCACCAAGATTTGGTGTGGAAGAAACAGGAAGCAATATCATTTAGCTTT 2928  
Db 1744 TATGTAAACAGCTAGCACCAAGATTTGGTGTGGAAGAAACAGGAAGCAATATCATTTAGCTTT 1803  
QY 2929 CAACACACAGAGGGAGCACCAAGCATCTCTCCAAAGATGTTTATTTATCAAACTCAGT 2988  
Db 1804 CAACACACAGAGGGAGCACCAAGCATCTCTCCAAAGATGTTTATTTATGCAAACTCAGT 1863  
QY 2989 TCTTCATCAATAATTTCTTTCTGGACACCTCTTCAAAACCTTAATGGGATTAACAATAT 3048  
Db 1864 TCTTCATCAATAATTTCTTTCTGGACACCTCTTCAAAACCTTAATGGGATTAACAATAT 1923  
QY 3049 TACTCTGTTTATTAACAGAAATACCTTCAGTACTTTTATGAGAAATTTTACCTCCATGAA 3108  
Db 1924 TACTCTGTTTATTAACAGAAATACCTTCAGTACTTTTATGAGAAATTTTACCTCCATGAA 1983  
QY 3109 CTAAACCAATGACTTTGACAAATAGTCTATCCCAATTTATAGATAAACTGACAAATTTTC 3168  
Db 1984 GTAAACCAATGACTTTGACAAATAGTCTATCCAAATTTATAGATAAACTGACAAATTTTC 2043  
QY 3169 AGCTACTATACATTTTGGTTAAACAGCAAGTACTTTCAGTTTGGAAATGGGAATAAAAGCAGT 3228  
Db 2044 AGCTACTATACATTTTGGTTAAACAGCAAGTACTTTCAGTTTGGAAATGGGAATAAAAGCAGT 2103  
QY 3229 GACATCATTTGAAGTATACAGATCAAGACATACCTGAAAGGTTTGTGTTGAAACCTGACT 3288  
Db 2104 GACATCATTTGAAGTATACAGATCAAGACATACCTGAAAGGTTTGTGTTGAAACCTGACT 2163  
QY 3289 TACGAATCCATTTCTGCTCACTGCATTAATGTAAAGTGGTCCCAACCGCTCAACCAAC 3348  
Db 2164 TACGAATCCATTTCTGCTCACTGCATTAATGTAAAGTGGTCCCAACCGCTCAACCAAC 2223

|    |      |  |      |  |    |      |   |      |  |
|----|------|--|------|--|----|------|---|------|--|
| QY | 3349 | GGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGCAGACTCTCCGCCATGTGAGACCA   | 3408 |  | Db | 3304 | CAGTCACTAGTGGCAACATTGACATGGATAAGACCTGACACTATCTCTGGCTACTTTCAA    | 3363 |  |
| Db | 2224 | GGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGCAGACTCTCCGCCATGTGAGACCA   | 2283 |  | QY | 4489 | AATTAACAAATTAACAACCTTCGTGCTCAAAAATGCAAAAGATGGGAATCCGGAAGAA      | 4548 |  |
| QY | 3409 | CCTCTTGTGTACATATGAGAGAGCATATATTTTGTGATAATCTGGAATAATACACTGATTAT | 3468 |  | Db | 3364 | AATTAACAAATTAACCACCTCAACTTCGTGCTCAAAAATGCAAAAGATGGGAATCCGGAAGAA | 3423 |  |
| Db | 2284 | CCTCTTGTGTACATATGAGAGAGCATATATTTTGTGATAATCTGGAATAATACACTGATTAT | 2343 |  | QY | 4549 | TGTGTTGATATCAAAAATTCAAATACCTCTATGAGCTCACTTAACCTGAAGAGACAGTA     | 4608 |  |
| QY | 3469 | ATATTAATAATTAATCTCCATCAACAGAAAGGGATTTCTGTGATACCTATATCTGCCAGCTA | 3528 |  | Db | 3424 | TGTGTTGATATCAAAAATTCAAATACCTCTATGAGCTCACTTAACCTGAAGAGACAGTA     | 3483 |  |
| Db | 2344 | ATATTAATAATTAATCTCCATCAACAGAAAGGGATTTCTGTGATACCTATATCTGCCAGCTA | 2403 |  | QY | 4609 | TATGGATTAAGAAAAATTTAGATGGTATAGATTTCAAGTGGCTGCCAGCACCAATGTGCGC   | 4668 |  |
| QY | 3529 | TACATCAAGACTGAAGAGATGTCACAGAAATCTTACCAATAATCAACACTTTTAAARAC    | 3588 |  | Db | 3484 | TATGGATTAAGAAAAATTTAGATGGTATAGATTTCAAGTGGCTGCCAGCACCAATGTGCGC   | 3543 |  |
| Db | 2404 | TACATCAAGACTGAAGAGATGTCACAGAAATCTTACCAATAATCAACACTTTTAAARAC    | 2463 |  | QY | 4669 | TATGGCAATGCTTTCAAACTGGATTTCTCAAAAACTCTGCCCTGGCCCTCCAGATGGTCTT   | 4728 |  |
| QY | 3589 | CTTTCTCTACCTCAGTCTCTTATCATGGGATCCCCAGTAAAGCAAAATGGTGAATA       | 3648 |  | Db | 3544 | TATGGCAATGCTTTCAAACTGGATTTCTCAAAAACTCTGCCCTGGCCCTCCAGATGGTCTT   | 3603 |  |
| Db | 2464 | CTTTCTCTACCTCAGTCTCTTATCATGGGATCCCCAGTAAAGCAAAATGGTGAATA       | 2523 |  | QY | 4729 | CCGTGAATGTTTATGTAGTAGTACCAATCACCTTTTAGCATCAGCATAGCTGGAGTGAA     | 4788 |  |
| QY | 3649 | ATAAGTTATGATTAATCTTACAGGACCAATAGAAAATTAATCTTTCAATTAATCTTGAT    | 3708 |  | Db | 3604 | CCGTGAATGTTTATGTAGTAGTACCAATCACCTTTTAGCATCAGCATAGCTGGAGTGAA     | 3663 |  |
| Db | 2524 | ATAAGTTATGATTAATCTTACAGGACCAATAGAAAATTAATCTTTCAATTAATCTTGAT    | 2583 |  | QY | 4789 | CCTGCTGTCAATTAATCTGAGCCAACTGTTATCTGTATGTATGTCAAAATCGGTAGATATGAT | 4848 |  |
| QY | 3709 | AATTACATAATATTGGAAGACTTTCACCAATTTACATATATAGCTTTTGTGCGCGCA      | 3768 |  | Db | 3664 | CCTGCTGTCAATTAATCTGAGCCAACTGTTATCTGTATGTATGTCAAAATCGGTAGATATGAT | 3723 |  |
| Db | 2584 | AATTACATAATATTGGAAGACTTTCACCAATTTACATATATAGCTTTTGTGCGCGCA      | 2643 |  | QY | 4849 | GAATTTAATATATCTTTCATCAAGTCAAAATGAAGAAAAATAAAACCATAGAAATTAAGAT   | 4908 |  |
| QY | 3769 | AGAACTAGAAAAAGCACTTGGTCTCTCCAGTATTTCTTTTCTTTTACAGATGAGTCAGTG   | 3828 |  | Db | 3724 | GAATTTAATATATCTTTCATCAAGTCAAAATGAAGAAAAATAAAACCATAGAAATTAAGAT   | 3783 |  |
| Db | 2644 | AGAACTAGAAAAAGCACTTGGTCTCTCCAGTATTTCTTTTCTTTTACAGATGAGTCAGTG   | 2703 |  | QY | 4909 | TTAGAAATTTTCAAGGATTTCTGTAGTGATCACTGCAATTTACTGGGAACATTAAGTCT     | 4968 |  |
| QY | 3829 | CGTTAGCACTTCCACAAATTTGACTTTTAATCAACTGACTGACTGACTGACTGACTGACTG  | 3888 |  | Db | 3784 | TTAGAAATTTTCAAGGATTTCTGTAGTGATCACTGCAATTTACTGGGAACATTAAGTCT     | 3843 |  |
| Db | 2704 | CGTTAGCACTTCCACAAATTTGACTTTAATCAACTGACTGACTGACTGACTGACTGACTG   | 2763 |  | QY | 4969 | GCATATGTAGAGAGGAAGTCAAGTCTGAAATGATTTGTGTACTTCTTGTAGAAATCAAGTCT  | 5028 |  |
| QY | 3889 | AAATGAGCCCAAGTCTCTCCAGGTGATTTGTTAAAGTATATAGTTTAAATTTCAAT       | 3948 |  | Db | 3844 | GCATATGTAGAGAGGAAGTCAAGTCTGAAATGATTTGTGTACTTCTTGTAGAAATCAAGTCT  | 3903 |  |
| Db | 2764 | AAATGAGCCCAAGTCTCTCTCCAGGTGATTTGTTAAAGTATATAGTTTAAATTTCAAT     | 2823 |  | QY | 5029 | AAGGACCCCACTTAAACAAATGATCACTTTAGAGAGATACAGATGAGTTTCAAAAATTTCAA  | 5088 |  |
| QY | 3949 | GAACATGAACCTGACACTATATATTAAGAAATATATCAGGATTTTAAATCTGAAGCCAAA   | 4008 |  | Db | 3904 | AAGGACCCCACTTAAACAAATGATCACTTTAGAGAGATACAGATGAGTTTCAAAAATTTCAA  | 3963 |  |
| Db | 2824 | GAACATGAACCTGACACTATATATTAAGAAATATATCAGGATTTTAAATCTGAAGCCAAA   | 2883 |  | QY | 5089 | TTAAGTCT    | 5148 |  |
| QY | 4009 | CTTGTGGACTGGAACAGTCAAGCACTTCTATCCGTGATCTCGGTTTCAACAAAGTT       | 4068 |  | Db | 3964 | TTAAGTCT    | 4023 |  |
| Db | 2884 | CTTGTGGACTGGAACAGTCAAGCACTTCTATCCGTGATCTCGGTTTCAACAAAGTT       | 2943 |  | QY | 5149 | TACCGAGAGAGATGATCT    | 5208 |  |
| QY | 4069 | GGAAATGGCAATCAATTTAGTAATGTAGTAAATTCACAAACCAAGATCAGTTCCAGAT     | 4128 |  | Db | 4024 | TACCGAGAGAGATGATCT    | 4083 |  |
| Db | 2944 | GGAAATGGCAATCAATTTAGTAATGTAGTAAATTCACAAACCAAGATCAGTTCCAGAT     | 3003 |  | QY | 5209 | AACACATTCGTCAATTTGCAATGTAGAGGACTTAAAGGTGGACATACATACATATCAGT     | 5268 |  |
| QY | 4129 | GTGCTGAGAAATATGAGTGCATGGCAACTAGTGGCAGTCAAGTTTGTAGTGAATGGGAT    | 4188 |  | Db | 4084 | AACACATTCGTCAATTTGCAATGTAGAGGACTTAAAGGTGGACATACATACATATCAGT     | 4143 |  |
| Db | 3004 | GTGCTGAGAAATATGAGTGCATGGCAACTAGTGGCAGTCAAGTTTGTAGTGAATGGGAT    | 3063 |  | QY | 5269 | GTTTAGCAGTCAATAGTGTGCTGAGGTTCCAAAGTTTCCGATGAGATTAACCATGGAT      | 5328 |  |
| QY | 4189 | CCACCCAAAGGCAATAGGAATATAACGAGTATATGGTAAACAGTTGAAGGAATTTCT      | 4248 |  | Db | 4144 | GTTTAGCAGTCAATAGTGTGCTGAGGTTCCAAAGTTTCCGATGAGATTAACCATGGAT      | 4203 |  |
| Db | 3064 | CCACCCAAAGGCAATAGGAATATAACGAGTATATGGTAAACAGTTGAAGGAATTTCT      | 3123 |  | QY | 5329 | ATCAAAGTCTCCAGCAGACCAAAAAACCAAAACCAACCAACCAACCAACCAACCAACCA     | 5388 |  |
| QY | 4249 | ACAAAGTTTCTCCCAAGATCAGTACACTTTTCAATAAGCTTCTTCCCAATCCTCA        | 4308 |  | Db | 4204 | ATCAAAGTCTCCAGCAGACCAAAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCA     | 4263 |  |
| Db | 3124 | ACAAAGTTTCTCCCAAGATCAGTACACTTTTCAATAAGCTTCTTCCCAATCCTCA        | 3183 |  | QY | 5389 | CTGCTTGTGACTTCAACAAATTAATTAATCAGAAATGCAATATGTTTACTACAGTATGAT    | 5448 |  |
| QY | 4309 | TATGCTTTTAAAGTAAAGCTTCAACCTCAGCTGGTGAAGGTGATGAAGACCATGCCAT     | 4368 |  | Db | 4264 | CTGCTTGTGACTTCAACAAATTAATTAATCAGAAATGCAATATGTTTACTACAGTATGAT    | 4323 |  |
| Db | 3184 | TATGCTTTTAAAGTAAAGCTTCAACCTCAGCTGGTGAAGGTGATGAAGACCATGCCAT     | 3243 |  | QY | 5449 | CATGACCAATTAATAATGCTTCAAGTGTTCGACAGAAACAGGAGCTCAGCATGATGGA      | 5508 |  |
| QY | 4369 | GTGAGCACTACTCTGAAACAGTTCCAGTGTTCACCAAAATATTCGCTTTTCTGATGTT     | 4428 |  | Db | 4324 | CATGACCAATTAATAATGCTTCAAGTGTTCGACAGAAACAGGAGCTCAGCATGATGGA      | 4383 |  |
| Db | 3244 | GTGAGCACTACTCTGAAACAGTTCCAGTGTTCACCAAAATATTCGCTTTTCTGATGTT     | 3303 |  | QY | 5509 | AATGTAAACAAAGTGTGATGATGATATTTTAAATAAGCAAGGCCATATTTTACAAATGAA    | 5568 |  |
| QY | 4429 | CAGTCAACTAGTGCACATTTGACATGGATGAACCTGACACTATCTCTGGCTACTTTCAA    | 4488 |  |    |      |   |      |  |



QYATYTRARAKPDSLEVLTLNLPNGTYEIKVAENNAGIGVSDPFLFOTARSAPGV  
VNLTVYALNSAVNLWYLPNGKITSFKLSVKHARSIGVVKDVSLEVEDILASGL  
PECNENESFLMSVTSSTPGLRVTPTVTTQSSSTAARSKISSWKEPISFWPTLIR  
PTTYLFEVASVTSAGYIDSTVTPESVPEGPPQNCIGNVTKAFSISWDPPTIV  
TKFSYRVELYSGRILDSNKTDLRFAFTHLTPFTMYDYVAAETSAGVGPKNLSV  
FTPPDVGAFDQLAEVATEIRITWRKPPNGIIISQYRVKVSLEVTGVLENTLL  
TQDESI SNMSPPEIMNLVDMGIFYEGSGEMSDLHSPASFIYNHPNDPASTRA  
EQSSPVTRNQMTDITAQLSVVPRVLPFTTEHTISVSAPTINGEGEPVTVLVRT  
RCQVPSIQIINKNISSSILLYWDPYNGKIHTHYIATELDTNKAQMTVDN  
FLITGLTKYIKYKORVAATHVGSSEENDI FVRTPEBPSPQDVQVTSVPS  
ELRLKWSPEKNGIIIAEVLVYQADTLFVNKTSTDIISDLKPYTLNIIISRYT  
LKGHNSOSSLVSRTSPDAPENITKNISGEIISFLPPRSPNGIIIOKTYIY  
LKRNSHEARTINTTSLQTIQGLKXYTHVLEVSASTLKGIRSRPISILTEBDAP  
DPPQNFVKQLSGVTVMLSWQPPLEPNGIILYTVYVWKSLSLRINATEASLVSD  
LDYNDYGACTASTRGDNKARSSINIRTPEGSPDPNDVHYVNLSSSIIILFW  
PPVKNIGIIQYYSVYQNTSGTFVQNTLLQVTKESDNTVVSARIYLAIFSYTFWL  
TASTSVGNKSDIIVHTDQIDPEGVNLTFSISSTAIHVSWEPPSQNGLVFY  
YLSNLQOQPPRHMIPLVTVENSIDPDLEKYTVI PKITPSTKGFSTVITOLHI  
KTEEDVPDTPPIINTFNKLSSTISLLSWDPPLKNGAILGYHLTQGPCHANFTVTS  
NHLVLELSPFTLYSFPAAARTMKGLPSSILFFYDESAPLAPPQNLTLINYSDFV  
MLTWSPLPGGLVIVKYSKIHEDTDIVYKNISGLQTDALGLEPVSITYSVSA  
FTKVGNGNYSNVETFTQESVPEAVRNECVARDQSVSVWRDPRKTNGLIIHYMI  
TVGNSTKYSVPRDPTFTFKLLPNTSVFEVRASTAGNESRCDISTLPETVSAP  
TNVAFNSVOSTSATLTWKPDITPGYFQNYKITTOLRAQCREWPEECIEHQKQYL  
YRANOTTEETHGLKFERWRFQVAASTNVGYSNASEWISTQTLPGPPDGPENHVVA  
TSFPGINISWSPAVITGTFYLLDVKSVDDDPNISFLKSNEENKTYEINLEVTR  
YSVITAFVGNVRAVDGKSAEVIITLESVPKDPNNMTFQKIPDEVTKFQTLFL  
PPSQNGNIRVYQALVREDDPTAVQIHNFSIIQKTDTSIIAMLEGLKGGHTYNISVY  
AINASGAVKQVMEITMDIKAPAPKSKPIPIRDATKLLVSTTTIIRMPICYND  
HGPINRVQVLAETQAOQDQNTKWDAYFNKARPYFTNVEGPPNPICIEGKTKFSND  
EVIYGDACNACI RGNEEKICNGPLKPKQYLFKFRATNMVQGFDSYSDPIKTLGE  
GLISERVEIILSVLILSILLLGTALFAPFIRQKQEGGYSYSPDAELIITKSKLD  
QLITVADLEKRLRLSLYSKIPISKSKFLQHVSELCTNSNLKFOESESLEPKF  
LDLQSLDADLPMNARFPNPKYNNRNLKLIADVSLPGSDYINASTVSYGLCPNE  
FIATQFLPGLTFDFWRVWETRTKTYLMLTQCFEKGIRKCHQYWPEDNKNPVYFGDI  
VITKLMEIDIQWITIRDLKIERHGDCTVRCQNTGWPHEGVPEPNTPLIHRKLVRT  
SRAHDTTPMVHSCAGVGTGFIALDHLTHOINNHDFVDIYGLVAELSRMCMVQN  
LAQYIFLHQCLDLLSNKGHQPVCFVNTSLQKXDSLDMESGDVELEWEETM"  
177. .230  
/genes="PTPRQ"  
177. .230  
/genes="PTPRQ"  
231. .7082  
/genes="PTPRQ"  
/product="glomerular mesangial cell receptor  
protein-tyrosine phosphatase"  
342. .638  
/genes="PTPRQ"  
/notes="fibronectin type III-like domain"  
639. .1088  
/genes="PTPRQ"  
/genes="PTPRQ"  
/note="fibronectin type III-like domain"  
1089. .1352  
/genes="PTPRQ"  
/notes="fibronectin type III-like domain"  
1353. .1622  
/genes="PTPRQ"  
/note="fibronectin type III-like domain"  
1623. .1868  
/genes="PTPRQ"  
/notes="fibronectin type III-like domain"  
1869. .2144  
/genes="PTPRQ"  
/note="fibronectin type III-like domain"  
2145. .2450  
/genes="PTPRQ"  
/notes="fibronectin type III-like domain"  
2451. .2732  
/genes="PTPRQ"  
/note="fibronectin type III-like domain"  
2733. .3014  
/genes="PTPRQ"  
/notes="fibronectin type III-like domain"  
3015. .3329  
/genes="PTPRQ"

sig\_peptide

mat\_peptide

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

/note="fibronectin type III-like domain"  
3330. .3623  
/genes="PTPRQ"  
/note="fibronectin type III-like domain"  
3624. .3899  
/genes="PTPRQ"  
/note="fibronectin type III-like domain"  
3900. .4184  
/genes="PTPRQ"  
/note="fibronectin type III-like domain"  
4185. .4523  
/genes="PTPRQ"  
/note="fibronectin type III-like domain"  
4524. .4787  
/genes="PTPRQ"  
/note="fibronectin type III-like domain"  
4788. .5099  
/genes="PTPRQ"  
/note="fibronectin type III-like domain"  
5100. .5399  
/genes="PTPRQ"  
/note="fibronectin type III-like domain"  
5400. .5864  
/genes="PTPRQ"  
/note="fibronectin type III-like domain"  
5901. .5969  
/genes="PTPRQ"  
/note="transmembrane-region site"  
6129. .6950  
/genes="PTPRQ"  
/note="protein-tyrosine phosphatase domain"

## ORIGIN

Query Match 74.5%; Score 5294.2; DB 10; Length 7396;

Best Local Similarity 84.7%; Pred. No. 0;

Matches 6048; Conservative 0; Mismatches 1049; Indels 45; Gaps 8;

QY 1 TAATTGTGTACTTGCAGAGGATCTGCTTTAAATCATTAATGACGGCAACATTTCTCT 60

Db 74 TAATTGTGTACTTGCAGGGGATCTGACTTTAAATCATTAATGACGGCAACATTTCTCT 133

QY 61 CTAGAGCCATCAATGTGATTCTACTGGCTG-AAAAATGTAATAAGATGGATTTCTTAT 119

Db 134 CCAGAGCCATCAATGTGATTGCTGTAGAAAAAATGTAATAATGATGATTTTCATTT 193

QY 120 CATTTTCTTTTACTTTTATTTGGGACTTCAGAGACACAGGTTGATGTTCCAAATGCT 179

Db 194 CTCCTTCTTTCTCTCTCATTTGGAACCTTCAGAGTCACAGGTTGATGTTTCCAGTTCTTT 253

QY 180 TCCTGTGTACTAGTAGATATAACCATCTCTTCAATTTCT---ACAACATACACCTCAC 236

Db 254 TGATGGCATCGGTATGATATACCGTATCTTCAGTTCTGCCACACCTACAGCTCGCC 313

QY 237 TGTACTAGATAGTGACACCAAAATGTAAACAAACCCAGGGCCTTCCAGTCTTCTAGCCGG 296

Db 314 TGTAGTAGAACACTGGCAACAAATGTGACAAACCCGGGSCCTTCCAGTCTTCTAGCCGG 373

QY 297 GGAAAGAGTCGGATCTGTGGGATTTCTTGTCTTGGAAATACACACCTTAATCCAAATGG 356

Db 374 GGAGAGAGTTGGCTCTGTCTGGGATACTCTCTTCTGTGGAAATACCCCTCTTAACCAATGG 433

QY 357 RAGGATATATCTTCAATTTGTCAAAATATAAGGAAGTTTCTCGTGTGATCAAAACATATA 416

Db 434 AAGAATCATATCTTATGTTGTTAAGTACAGGAAGTCTGTCCATGATGACAGACATA 493

QY 417 TACACAAGTCAGATCAAAAGCCAGACAGTCTGGAAGTTCTTCTTAATAATCTTAACTGG 476

Db 494 TACCCGAGCGAGAGCGAAGCCAGACAGTCTGGAGGTTCTTCTCACTAACTTAATCTGG 553

QY 477 AACACATATGAAATTAAGTTGCTGCTGAAAAACAGTGTGGCATTTGGAGTGTTTAGTGA 536

Db 554 AACACATATGAAATTAAGTTGCTGCTGAAAAACATGCTGGCATCGGAGTGTTTAGTGA 613

QY 537 TCCATTTCTCTTCCAACTGCAGAAAGTGTCTCCAGGAAAGTGGTGAATCTCACAGTTGA 596  
DB 614 CCCATTTCTTTTCAAACTGCAGAAAGTGTCTCCGGGTAAAGTAGTTAACTCTCACAGTGA 673  
QY 597 GGCTCAACAGCTTTCAGCAGTAAAGTGTATTTGGTATTTTAACTTCGGCAACAAATGGCAA 656  
DB 674 AGCTCTCAACTACTCTGCAGTAAACTTGTATTTGGTATTTTAACTTCGGCAACAAATGGCAA 733  
QY 657 AATTACAGCTTCAAGATTAAGTGTCAAAATGCGAGAAAGTGGGATAGTAGTGAAGATGT 716  
DB 734 AATCACCAGCTTCAAGATTAAGTGTCAAGCAACGCGAGAAAGTGGGATAGTAGTCAAGATGT 793  
QY 717 CTCATCAGAGTAGAGGACATTTTGTCTGGGAAATTCAGAGATGCAATGAGATAGTGA 776  
DB 794 CTCCTCAGAGTAGAGGACATTTTGTCTGGGAAATTTAACTGAAATGTAATGAGAAACGCGA 853  
QY 777 ATCTTTTATGAGTAGACGACGCTTCTCAACCTTGGTAGAGTTACACCTCCATC 836  
DB 854 ATCTTTTGTGGATGACCAAGGCTTCTCCAACTGGTAGAGTGACACTACGT 913  
QY 837 GCGTACCAACATTTCAATCAAGCAGCTTGAACAGAAATGAGATCAGCTCTGTGTGGAAGA 896  
DB 914 TCGCACCAACAGCTATCCAGCAGCGCGCAAGAGTAAAGTCAAGCTCTGTGTGGAAGA 973  
QY 897 GCCTATCAGTTTGTAGTAGACACTTGAACACCTTATACAAATATCTTTTGAAGTTTC 956  
DB 974 GCCCATCAGTTTGTGTGTGACTCACTTTGAGACCTTACACAACTATCTCTTTGAACTCTC 1033  
QY 957 AGCTGTACAACTGAAGCAGTTATATGATAGTACGATTTGTCAGAACCAACCAATCAGT 1016  
DB 1034 AGCCGTACACCGAAGCAGATATATGACAGCACCATTTGTAGGACGCCAGATCAGT 1093  
QY 1017 GCCTGAAGGACCAACCAACTCGGTAAACAGGCAACATCACAGGAAAGTCTTTTCAAT 1076  
DB 1094 GCCTGAAGGACCAACCAACTGTATATGGGGAATGTACAGGAAAGGCTTTTCAAT 1153  
QY 1077 TTTATGGGACCAACCAACTATAGTAACAGGGAATTTAGTTATAGAGTTGAATATATGG 1136  
DB 1154 TTCTGGGACCAACCAACTATTTGTACAGGAAAGTTTGTATACAGAGTTGAATTTGTATGG 1213  
QY 1137 ACCATCAGGTGCGATTTTGGATTAACAGCAACCAAGACCTCAAGTTTGCATTTCACTAACCT 1196  
DB 1214 GCCTCAGGCGGATTTTGGATTAATAGCAGAAAGACCTCCGGTTTGGCTTTTACATCT 1273  
QY 1197 AACACATTTTCAATGTATGATGTCTATATTTGGGCTGAAACAGTGCAGGGGACTGGGCC 1256  
DB 1274 CACACCATTCACAAATGTACGATGTCTATGTCTGCGGAAACCAAGTCTGGGTTGGACC 1333  
QY 1257 CAAGTCAAAATTTTCACTCTCAACAGATGTTCAGGGGCAAGTGTTCATTTACA 1316  
DB 1334 CAAGTCAAAATCTTTCAAGTATTTACTCTCCCAAGATGTTCAGGAGCAGTGTTCATTTACA 1393  
QY 1317 ACTTGCAGAGGTAGATTCACGCAAGTAAAGATTTCTTGGAGAAACCAAGCAACCAAA 1376  
DB 1394 GATCGCAGAGGTAGAGGCAACAGAAATAAGATTTCTTGGAGAAACCTCGACAAACAA 1453  
QY 1377 TGGAAATTTTAAACCAATACCGAGTGAAGTGTAGTTTCAGAGACAGGGAATTAATTTTGA 1436  
DB 1454 CGGAATCATCAGCCAGTACCGTGTGAAGTGTCAAGTCTCTAGAGACAGGAGTAGTTTGA 1513  
QY 1437 AAATACCTTTGCTCACTGGAAATTAAGTATATAAATGACCCCATGGCTCCAGAAATTTG 1496  
DB 1514 AAATACCTGCTCACAGGACCAAGATGAGTCTAATAGCAACCCCATGTCTCCAGAAATTA 1573  
QY 1497 GAACATAGTAGCCCAATGTTAGGATTTATAGGGTTTACAGGATGATGTCTGACCT 1556  
DB 1574 GAACCTAGTAGACCCCATGATAGGTTCTATAGGGCTCAGGAGAAATGTCTTCGACCT 1633  
QY 1557 TCACTCACTTTGTACATTTATATAAAGCCCATCAGATATAAAATTTCTCTCAAGGAA 1616  
DB 1634 GCATCCCGGCTTCAATTCATATACCAAGCCCATCACAATGACTTTCTCTGCCAGCAC 1693  
QY 1617 TAGAGCTGAAGACCAAGTCTTCCAGTTGTAACTACAGGAATCAGTATATTAATCTGACAT 1676

DB 1694 TAGAGCTGAAGACAAAGTTCCAGTTGTAAACCAAGGAATCAGTATATATGACCGCAT 1753  
QY 1677 TGCAGCTGAACAGCTGTCTTATGTTATCAGGAGACTTGTACCTTTCTACTGAGCACATGAT 1736  
DB 1754 CACTGCCGAACAGCTGTCTTACGTTGTCAAGCGGTTGTACCAATCTACTGAGCACAGAT 1813  
QY 1737 TAGTGTATCTGCTTTTCAACATCATGAGGAGAGGACCAACCAAGTGTCTCAGTGTTAGGAC 1796  
DB 1814 TAGCGTATCTGCTTTTCAACATCATGAGGAGAGGACCAACCAAGTGTCTTACTGTAGGAC 1873  
QY 1797 ACCTCAGCAGAGTCCAGCTCCATTAATAATTAATAACTATAAAATATTAAGTTCTTCATC 1856  
DB 1874 ACCTGAGCAGAGTCCAGCTCTATCCAAATTAATAACTATAAAATATTAAGTTCTTCATC 1933  
QY 1857 TATTTTGTATATTTGGGATCTCCAGATATCCAAATGGAATAATACTCACTATACGAT 1916  
DB 1934 TATTTTACTGTATTTGGGATCTCCAGATATCCAAATGGAAGATCACTCATTTACACAT 1993  
QY 1917 TTAATCAATGGAATTTGGATACAAACAGAGCATTTCCAGATAACTCAATAGATAACAGCTT 1976  
DB 1994 CTACGCAACGAGCTGGAACACAAACAGAGCATTTCCAGATGACCACTGTAGATAACAGCTT 2053  
QY 1977 TCTCATAAACGAGTTTAAAGAAATACAAAAATACAAATGAGAGTGGCAGCTCAACCCA 2036  
DB 2054 TCTCATACAGGTTTGAAGAAATATACAAAGATCAAAATGAGGTTGTGCTCAACTCA 2113  
QY 2037 CGATGAGAGAAAGTTCTTGTCTGAAGAAATGACATCTTTGTGAGAACTTTCAGAAATGA 2096  
DB 2114 TGTGTGGGAAAGTTCTTGTCTGAGAAATGACATCTTTGTGAGAACTTTCAGAAATGA 2173  
QY 2097 ACCGGAATCATCTCAAGATGTTCGAAGTAAATGATTTGTACCGAGATGAAATAAGGTT 2156  
DB 2174 GCCAGATCATCACTCAAGATGTTCAGTAACTGTGTCTTCCAAAGTGAACCTGAGATT 2233  
QY 2157 GAAGTGTCAACACCGGAAAGCCCAATGGGATCATTTTGTCTTATGAGTGTCTATATA 2216  
DB 2234 GAAGTGTCTCCCGCTGAAAGAACCTTAAACGGGATCATTTTGTCTTATGAGTGTCTTACCA 2293  
QY 2217 AAATATAGATATCTTTATATATGAAGAAACATCAACCAACAGACATAATATAAGGAAT 2276  
DB 2294 AAATGACAGACCTTGTGTAAAGAAACCTCAACACAGACATCATTTATAGTACCT 2353  
QY 2277 AAGACCTCAACCTCTATAACATTTCTGTAAAGTCTTACACAGATTTGGTCAATGCAA 2336  
DB 2354 GAAACCTTACACACTCTATAATTTCTATCCGGTCTGTATACAGACTTGGCCATGGCAA 2413  
QY 2337 TCAGTGTCTTTTACTCTCTGTAAAGACTTCCGAGACTGTGCTGTATAGTGCACAGA 2396  
DB 2414 TCAGTGTCTCTTCTGTCTGTAAAGACTTTCAGAGACTGTGCTGTATAGTGCACAGA 2473  
QY 2397 AAATATCACTTACAAATAATTTTCTTGTGAGAGATTTGAGCTATCATTTCTTCCCAAG 2456  
DB 2474 AAATATCTTACAAATAATTTTCTTGTGAGAGATTTGAGATATCTTCTTCCCAAG 2533  
QY 2457 TAGTCCCAATGGAATCATTAATAATAATATATCTCAAGAGAAAGTAATGGAATGA 2516  
DB 2534 GAGTCCCAACGCAATCATACAAATAACAAATTTTACTCAAGAGAAAGTAATAGCCATGA 2593  
QY 2517 GGAAGAACTATATAATACAACTTTTAAACCAAAACATTAAGTACTGAAGAAATATAC 2576  
DB 2594 GGCAGAACTATAATAATACAACTTTTGAACCAACCAATTTGGAGGACTGGAAGATATAC 2653  
QY 2577 CCATATATCATTTGAGTGTCTGTAGTACACTTGAAGGTTGAGGAGTTCGGAGTGTCTC 2636  
DB 2654 CCATATGTAATCGAGTGTCTGTAGTACACTTGAAGGTTGAGGAGTTCGGAGTGTCTC 2713  
QY 2637 CATAGTATACAGCGGAGGAGATGCTCTGATTTCTCCCTCAAGACTTCTCTGTAAA 2696  
DB 2714 AATAGCATATTTGACCGAGGAGAGCGTCTGATTTCTCTCTCAGAACTTCTCTGTGA 2773  
QY 2697 ACAGTGTCTGTGTCAACGTTGATGTGTCAATGGCAACCAACCTCTGAGGCAATGGAAT 2756

Db 2774 ACAGTTGTCTGGTGTACCGTGTATGTGTATGCGCAACACCCCTTGGAGCAAAATGGAAT 2833  
QY 2757 TATCCTTTTATTACAGAGTTTATGTCTGGATAGATCATCATTTAAACACTTATTAATGTAC 2816  
Db 2834 TATCCTCTATTACAGAGTTTATGTCTGGATAGATCATCATTTAAACACTTATTAATGTAC 2893  
QY 2817 TGAACATCATTTGGAGTTATCAGATTGGATTATTAATGTGAATACAGTCTTATGTAAC 2876  
Db 2894 TGAAGCATCACTGGTGTGTCTGACTTAGACTATATATGTGACTAGCGTCCGTGTAAAC 2953  
QY 2877 AGCTAGCACAGATTGGTGTATGGGAAACAGGAACAATATCATTTAGCTTTCAACACC 2936  
Db 2954 AGCTAGCACAGATTGGGATGGGAAACAGGAAGCAGCATCATTTAACTTTTGAACACC 3013  
QY 2937 AGGGAGACACCAAGCGATCTCTCCAAAGATGTTTATTATGCAACCTCAGTCTTCTATC 2996  
Db 3014 AGGGAGAGACCCAGTGATCTCTCCAAAGATGTTTATTATGCAACCTCAGTCTTCTATC 3073  
QY 2997 AATAATTTCTTTCTGGACACCTCTTTCAAAACCTAATGGATTATACAATATTACTCTGT 3056  
Db 3074 GATTAATTTCTTTTGGACTCTCTCTGTAAACCTAATGGATTATACAATATTACTCTGT 3133  
QY 3057 TTATTAACAGAAATPACTTCAGGTACTTTTATGCAAGAAATTTTACATCCATGAACCTAACCAA 3116  
Db 3134 TTATTAACAAATPACTTCAGGTACTTTTGTGCAAGAACTTTTACATCTTTCAAGTAACCAA 3193  
QY 3117 TGAATTTGCAATATGACTGTATCCCAATATTAGATAAATGCAATATTACGTACTTA 3176  
Db 3194 GGAATCTGCAATGTGACAGTGTCTGCAAGGATATACAGACTGGCAATATTACGTACTTA 3253  
QY 3177 TACATTTTGGTTAAACAGCAAGTACTTTCAGTTGGAATGGGAATPAAAGCAGTGACATCAT 3236  
Db 3254 TACATTTTGGTTAAACAGCAAGTACTTTCAGTTGGAATGGGAATPAAAGCAGTGACATCAT 3313  
QY 3237 TGAAGTATACAGATCAAGACATCTGAAAGGTTTGTGGAAACCTGACTTTACGAATC 3296  
Db 3314 TCATGTATACAGACCAAGACATACCTGAAAGTCTCTGTTGGAACCTGACTTTTGAGTC 3373  
QY 3297 CATTTGCTGACATGCAATTAATGTAAAGTGGTCCACCGGCTCAACCAACCGTCTAGT 3356  
Db 3374 CATTTTCACTGCAATTCATGTGAGCTGGGAGCCACCATCCCAACCAATGGCCCTGGT 3433  
QY 3357 CTTCATCTACTGTTTCTACTGATCTTACAGCAGA--CTCCTCGCATGTGAGACCACTCT 3413  
Db 3434 CTTCATCTACTGTTTCTACTGATCTTACAGCAGA--CTCCTCGCATGTGAGACCACTCT 3493  
QY 3414 TGTACATATGAGAGAGCATATTTTGTATATCTGGAATAATACATGATTATATAT 3473  
Db 3494 GGTACCTATGAAACAGCATAGACTTTGTATGATCTGGAATAATATACATGATTATAT 3553  
QY 3474 AAAAAATCTCCATCAACAGAAAGGATCTCTGATACCTATACCTACTGCCAGCTATACAT 3533  
Db 3554 TAAATCACTCCGTCAACAGAAAGGATCTCTGAGACCTATACCTACCCAACTACACAT 3613  
QY 3534 CAAGACTGAAGAGATGTCACAGAACTTCCACCAATTAATCAACACTTTTAAAAACCTTTC 3593  
Db 3614 CAACAGAGAGAGAGCTTCCAGACACTCCACCAATTAATCAACACTTTTAAAAATCTTTC 3673  
QY 3594 CTCTACCTCAGTCTCTTATCATGGGATCCCCAGATAAAGCCAAATGGTGCAATATAAG 3653  
Db 3674 TTCTACCTCAATCTCTTATCATGGGATCCCCCAATTAAGCCAAACGGTGCAATATAAG 3733  
QY 3654 TTATGATTTTAACTTTTACAGGACCAATGAAATTTATCTTCAATTTACTTTCTGATTAATTA 3713  
Db 3734 TTACCAATTTAACTTTTACAGGACCCCAAGCAATCTTCTGCTACTTCCGGTAAACCA 3793  
QY 3714 CATAATTTGGAAGAGCTTTCACATTTTACATTTATATAGCTTTTGTGCGCCCAAGAAC 3773  
Db 3794 CATAGTACTGGAAGAGACTCTCGCAATTTACCTTTATATAGTTTTTTTGTGCTGCAAGAAC 3853  
QY 3774 TAGAAAAGGACTGTGTCCTTCCAGTATCTTTTCTTTTACAGAGTGAAGTCAAGTCCGTT 3833  
Db 3854 AATGAAAGGACTCGGTCTTCCAGTATCTTTTCTTTTACAGAGTGAAGTCAAGTCCGTT 3913

QY 3834 AGCACTCCACAAAATTTGACTTTTAACTCAACTGTACTTTTCCAGACTTTTGTATGCTGAAATG 3893  
Db 3914 AGCACTCCCAAGATTTGACTTTTAACTCAACTGTACTTTTCCAGACTTTTGTATGCTGAAATG 3973  
QY 3894 GAGCCCAAGTCTCTTCCAGGTGGTATTTGTTAAAGTATATAGTTTAAATTTTCAATGAA 3953  
Db 3974 GAGCCCAAGTCTCTTCCAGGTGGTATTTGTTAAAGTATATAGTTTAAATTTTCAATGAA 4033  
QY 3954 TGAATCTGACATATATATTATAAGAAATATATCAGGATTTAAATTTTCAATGAAATTTTGT 4013  
Db 4034 TGAGACTGATATCTGATTTTATAAGAAATATTTTCAAGTCTTCCAGACTGATGCAAACTTGA 4093  
QY 4014 TGGACTTGAACCAAGTCAAGCAGCTCTATCCGTGTATCTGCTGCTGCTGCTGCTGCTGCTG 4073  
Db 4094 GGGACTTGAAGCTGTTAGCAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4153  
QY 4074 TGGCAATCAATTTAGTAAATTTTCAAAATTTCAAAACCCCAAGAAATCAGTTTCCAGATGTCGT 4133  
Db 4154 TGGCAATCAGTATAGTAAATTTTCAAAATTTTCAAAACCCCAAGAAATCAGTTTCCAGATGTCGT 4213  
QY 4134 GAGAAATGCAAGTGGCAATCTAGCTGGCAGTCAAGTTTGTAGTGAATTTGGAATTTCCACC 4193  
Db 4214 ACGGAATATAGAGTGGTGGCAAGTCAAGTGGCAGTCAAGTTTGTAGTGAATTTGGAATTT 4273  
QY 4194 CAABAAGGCAATGGAATTAATAACGAGTATATGTTAAAGTGGTGAAGAAATTTCTACAAA 4253  
Db 4274 GAGAAATCAATTTGGAATTAATAACGAGTATATGTTAAAGTGGTGAAGAAATTTCTACAAA 4333  
QY 4254 AGTTTCTCCCAAGATCAATGTACATTTTCAAAAGTCTTCTTGCAATTTACCTCATATGT 4313  
Db 4334 AGTCTCTCCCAAGATCAATGTACATTTTCAAAAGTCTTCTTGCAATTTACCTCATATGT 4393  
QY 4314 CTTTAAAGTAAAGCTTTCAACCTCAGCTGGTGAAGTGAAGAAAGCAATGCGCATGTGAC 4373  
Db 4394 CTTTAAAGTAAAGCTTTCAACCTCAGCTGGTGAAGTGAAGAAAGCAATGCGCATGTGAC 4453  
QY 4374 CACATCTCTGAAAGAGTTCCTGAGTTCCTGCAATTTTCTGATGTTTCTGATGTTTCTGATG 4433  
Db 4454 CACATCTCTGAGAGAGTTCCTGAGTTCCTGCAATTTTCTGATGTTTCTGATGTTTCTGATG 4513  
QY 4434 AACTAGTGAACATTTGACATGGAATGAAGCTGACATCTCTTGGCTACTTTTCAAAATTA 4493  
Db 4514 AACTAGTGAACATTTGACATGGAATGAAGCTGACATCTCTTGGCTACTTTTCAAAATTA 4573  
QY 4494 GAAATTTACCTCAACTTCTGCTGCTCAAAATTTGCAAGAAATGGGAATCCGAGAAATGCT 4553  
Db 4574 CAAGATAACTTCAACTTCTGCTGCTCAAAATTTGCAAGAAATGGGAATCCGAGAAATGCT 4633  
QY 4554 TGAATATCAAAATTTCAATCTCTTGAAGCTCACTTAACTGAAGAGACAGTATATGG 4613  
Db 4634 TGAATATCAAAATTTCAATCTCTTGAAGCTCACTTAACTGAAGAGACAGTATATGG 4693  
QY 4614 ATTAAGAAATTTAGATGTTATAGATTTCAAGTGGTGGCTGCGACCAATTTGCTGGCTATGG 4673  
Db 4694 GTTAAGAAATTTAGATGTTATAGATTTCAAGTGGTGGCTGCGACCAATTTGCTGGCTATGG 4753  
QY 4674 CAATGCTTCAAACTGGAATTTTCAAAATTTTCAAGTGGTGGCTGCGACCAATTTGCTGGCTATGG 4733  
Db 4754 TAAATGCTCGGAATGGAATTTTCAAAATTTTCAAGTGGTGGCTGCGACCAATTTGCTGGCTATGG 4813  
QY 4734 AATGCTTCAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4793  
Db 4814 AATGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 4873  
QY 4794 TGTCTACTTGGACCAACATGTTTCTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATG 4853  
Db 4874 TGTCTACTTGGACCAACATGTTTCTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATG 4933  
QY 4854 TAAATATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4913  
Db 4934 TAAATATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4993







|                            |  |  |      |        |                 |   |      |
|----------------------------|--|--|------|--------|-----------------|---|------|
| Db                         | 7150   | TCTTGCTTCTGAATGGAGCTGGGCA-----AAGGAAATCTTTATGCTTCCGGACCATG     | 7204 | Db     | 633             | GTCAAGTGTGAAATGATGTTTACTACTTTTAGAATCAGCCCCCAAGGACCCACCTAACAA    | 692  |
| Qy                         | 7105   | CT 7106  |      | Qy     | 5046            | CATGACATTTTCAAGAGATACACAGATGAGCTTACAAAATTTTCAATTAACGTTCTTCTCTCC | 5105 |
| Db                         | 7205   | CT 7206  |      | Db     | 693             | CATGACATTTTCAAGAGATACACAGATGAGCTTACAAAATTTTCAATTAACGTTCTTCTCTCC | 752  |
| RESULT 8                   |  |  |      | Qy     | 5106            | TTCTCAACCTAATGGAATATATCAAGTATATCAAGCTCTGGTTTACCGAAGAGATGATCC    | 5165 |
| AR073855                   |  |  |      | Db     | 753             | TTCTCAACCTAATGGAATATATCAAGTATATCAAGCTCTGGTTTACCGAAGAGATGATCC    | 812  |
| LOCUS                      | AR073855   | 3973 bp  | DNA  | linear | PAT 28-AUG-2000 |   |      |
| DEFINITION                 | Sequence 21 from patent US 5952212.                    |  |      | Qy     | 5166            | TACTGTCTGTCAGATTCACAACTCAGTATATATACAGAAAACCAACACATTCGTCAATGC    | 5225 |
| ACCESSION                  | AR073855   |  |      | Db     | 813             | TACTGTCTGTCAGATTCACAACTCAGTATATATACAGAAAACCAACACATTCGTCAATGC    | 872  |
| VERSION                    | AR073855.1   | GI:10000615  |      | Qy     | 5226            | TAATGTAGAGGACTTAAAGGTGGACATACATACATATCAGTGTGTTACGCAAGTCAATAG    | 5285 |
| KEYWORDS                   | Unknown.   |  |      | Db     | 873             | TAATGTAGAGGACTTAAAGGTGGACATACATACATATCAGTGTGTTACGCAAGTCAATAG    | 932  |
| SOURCE                     | Unknown.   |  |      | Qy     | 5286            | TGCTGTGTCAGGTCCAAAGGTTCCGATGAGAAATACCATGATATCAAAAGCTCCAGCAG     | 5345 |
| ORGANISM                   | Unclassified.  |  |      | Db     | 933             | TGCTGTGTCAGGTCCAAAGGTTCCGATGAGAAATACCATGATATCAAAAGCTCCAGCAG     | 992  |
| REFERENCE                  | 1 (bases 1 to 3973)                                    |  |      | Qy     | 5346            | ACCAAAAACCAAAACCAACCCCTATTTATGATGCCACAGAAAACCTGCTTGTGACTTCAAC   | 5405 |
| AUTHORS                    | Moller,N.Peter,Hundahl., Moller,K.Bach. and Ullrich,A. |  |      | Db     | 993             | ACCAAAAACCAAAACCAACCCCTATTTATGATGCCACAGAAAACCTGCTTGTGACTTCAAC   | 1052 |
| TITLE                      | Protein tyrosine phosphatase                           |  |      | Qy     | 5406            | AAACAATTACAATCAGAAATCCCAATATGTTTACTACAGTATGATCATGAGCAATAAAAA    | 5465 |
| JOURNAL                    | Patent: US 5952212-A 21 14-SEP-1999;                   |  |      | Db     | 1053            | AAACAATTACAATCAGAAATCCCAATATGTTTACTACAGTATGATCATGAGCAATAAAAA    | 1112 |
| FEATURES                   | Location/Qualifiers                                    |  |      | Qy     | 5466            | TGTAACAAGTCTTGGCACAGAAACAGGAGCTCAGCATGATGGAATGTAAACAAGTGGTA     | 5525 |
| source                     | 1..3973  |  |      | Db     | 1113            | TGTAACAAGTCTTGGCACAGAAACAGGAGCTCAGCATGATGGAATGTAAACAAGTGGTA     | 1172 |
| ORIGIN                     | /organism="unknown"                                    |  |      | Qy     | 5526            | TGATGATATTTTAAATAAGCAAGGCCATATTTTACAATAGAGGCTTCTTAACCTCC        | 5585 |
| Query Match                | 36.8%; Score 2618.6; DB 6; Length 3973;                |  |      | Db     | 1173            | TGATGATATTTTAAATAAGCAAGGCCATATTTTACAATAGAGGCTTCTTAACCTCC        | 1232 |
| Best Local Similarity      | 97.2%; Pred. No. 0;                                    |  |      | Qy     | 5586            | ATGTACAGAAGCAAGACAAAAGTTAGTGGCAATGAAGAAATCTACATCATAGGTGCTGA     | 5645 |
| Matches 2716; Conservative | 0; Mismatches 4; Indels 75; Gaps 2;                    |  |      | Db     | 1233            | ATGTACAGAAGCAAGACAAAAGTTAGTGGCAATGAAGAAATCTACATCATAGGTGCTGA     | 1292 |
| Qy                         | 4386   | AACAGTTCACAGTGTCCCAACAATATGCTTTTCTGATGTTTCAGTCAACTAGTCAAC      | 4445 | Qy     | 5646            | TAATGATCATGATTTCTTGGCAATGAAGCAAAAATTTGCAATGGAACCACTGAACCAAA     | 5705 |
| Db                         | 33   | AAAAGTTCACAGTGTCCCAACAATATGCTTTTCTGATGTTTCAGTCAACTAGTCAAC      | 92   | Db     | 1293            | TAATGATCATGATTTCTTGGCAATGAAGCAAAAATTTGCAATGGAACCACTGAACCAAA     | 1352 |
| Qy                         | 4446   | ATTGACATGGATGAAGACCTGACACTATCTTGGCTACTTTTCAAAATTTACAAAATTACCAC | 4505 | Qy     | 5706            | AAAGCAATCTTATTTAAATTTAGAGCTACAAATATTTAGGACAAATTTTACTGACTCTGA    | 5765 |
| Db                         | 93   | ATTGACATGGATGAAGACCTGACACTATCTTGGCTACTTTTCAAAATTTACAAAATTACCAC | 152  | Db     | 1353            | AAAGCAATCTTATTTAAATTTAGAGCTACAAATATTTAGGACAAATTTTACTGACTCTGA    | 1412 |
| Qy                         | 4506   | TCAACTTCGTGCTCAAAAATGCAAGATGGGAATCGAAGATGTGTTGAAATATCAAAA      | 4565 | Qy     | 5766            | TTATTCAGACCTGTTTAAAGACTTTAGGGGAAGGACTTTTCAGAAAAGAACCGTAGAGATCAT | 5825 |
| Db                         | 153  | TCAACTTCGTGCTCAAAAATGCAAGATGGGAATCGAAGATGTGTTGAAATATCAAAA      | 212  | Db     | 1413            | TTATTCAGACCTGTTTAAAGACTTTAGGGGAAGGACTTTTCAGAAAAGAACCGTAGAGATCAT | 1472 |
| Qy                         | 4566   | AATTCATACCTCTATGAAGCTCAGTTTAACTGAAGAGACAGTATATGGAATTAAGAAAT    | 4625 | Qy     | 5826            | TCTTTCCGTCACCTTTGTGATCTTTTCAATAATTTCTTGGAAACAGCTATTTTTCGAT      | 5885 |
| Db                         | 213  | AATTCATACCTCTATGAAGCTCAGTTTAACTGAAGAGACAGTATATGGAATTAAGAAAT    | 272  | Db     | 1473            | TCTTTCCGTCACCTTTGTGATCTTTTCAATAATTTCTTGGAAACAGCTATTTTTCGAT      | 1532 |
| Qy                         | 4626   | TAGATGGTATAGATTCCAAGTGGCTGCCAGCAACAATGCTGGCTATGGCAATGCTTCAAA   | 4685 | Qy     | 5886            | TGCAAGAAATTCGACAGAACGAGAAGAGTGGACATCTCTCTCAGAGTGCAGAAAT         | 5945 |
| Db                         | 273  | TAGATGGTATAGATTCCAAGTGGCTGCCAGCAACAATGCTGGCTATGGCAATGCTTCAAA   | 332  | Db     | 1533            | TGCAAGAAATTCGACAGAACGAGAAGAGTGGACATCTCTCTCAGAGTGCAGAAAT         | 1592 |
| Qy                         | 4686   | CTGGATTTCTACAAAACCTGCTGCTGCTCCAGATGCTGCTCCTGAAAATGTTTCATGT     | 4745 | Qy     | 5946            | TATTGACACTTAAATTCGAAGCTGATCAGTCTCATACAGTGGCAGACCTGGAACCTGAAGGA  | 6005 |
| Db                         | 333  | CTGGATTTCTACAAAACCTGCTGCTGCTCCAGATGCTGCTCCTGAAAATGTTTCATGT     | 392  | Db     | 1593            | TATTGACACTTAAATTCGAAGCTGATCAGTCTCATACAGTGGCAGACCTGGAACCTGAAGGA  | 1652 |
| Qy                         | 4746   | AGTAGCAACATCACCTTTTAGCATCAGCAATGAGTGAACCTGCTGCTCAATCTG         | 4805 | Qy     | 6006            | CGAGAGATTAACGCG-----GCGAATTAACGCG-----                          | 6020 |
| Db                         | 393  | AGTAGCAACATCACCTTTTAGCATCAGCAATGAGTGAACCTGCTGCTCAATCTG         | 452  | Db     | 1653            | CGAGAGATTAACGCGTACTTCTTCTTTTAGCGAAGGAGATTTTGTGTCATCCA           | 1712 |
| Qy                         | 4806   | ACCAACATGTTATCTGATGATGTCAAATCGGTAGATAATGATGAATTTAATAATATCTTT   | 4865 | Qy     | 6021            | -----GCCAATTAAGCAAGAAATCTCTCTGCAACATGT                          | 6053 |
| Db                         | 453  | ACCAACATGTTATCTGATGATGTCAAATCGGTAGATAATGATGAATTTAATAATATCTTT   | 512  | Db     | 1713            | GTACTTAGTTATAGAAAATCCATCAAGCCAAATAGCAAGAAATCTCTCTGCAACATGT      | 1772 |
| Qy                         | 4866   | CATCAAGTCAATGAAGAAATTAACCATAGAAATTAAGATTTAGAAATATTCACAG        | 4925 |        |                 |   |      |
| Db                         | 513  | CATCAAGTCAATGAAGAAATTAACCATAGAAATTAAGATTTAGAAATATTCACAG        | 572  |        |                 |   |      |
| Qy                         | 4926   | GTATTCGTAGTATCATCTGATTTTACTGGGAACATTTAGTGTGCTATGTAGAGGGA       | 4985 |        |                 |   |      |
| Db                         | 573  | GTATTCGTAGTATCATCTGATTTTACTGGGAACATTTAGTGTGCTATGTAGAGGGA       | 632  |        |                 |   |      |
| Qy                         | 4986   | GTCAAGTGTGAAATGATGTTTACTTCTTTAGAAATCAGCCCCCAAGGACCCACTAACAA    | 5045 |        |                 |   |      |





| VERSION                    | AX686451.1  | GI:29372156  |
|----------------------------|---|--|
| KEYWORDS                   | Homo sapiens (human)  |  |
| SOURCE                     | Homo sapiens  |  |
| ORGANISM                   | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  |  |
| REFERENCE                  | 1   |  |
| AUTHORS                    | Edinger, S., Macdougall, J.R., Millet, I., Ellerman, K., Stone, D.J., Gerlach, V., Grosse, W.M., Alsobrook, J.P., Lepley, D.M., Rieger, D., Burgess, C.E., Casman, S.J., Spytek, K.A., Boldog, F.L., Li, L., Padigaru, M., Mishra, V., Patturajan, M., Shenoy, S., Raetelli, L., Tchernev, V.T., Vernet, C.A., Zehusen, B.D., Malyankar, U.M., Guo, Y., Miller, C.E. and Gangolli, E.A. |  |
| TITLE                      | Proteins and nucleic acids encoding same  |  |
| JOURNAL                    | Patent: WO 02057450-A 7 25-JUL-2002;  |  |
| FEATURES                   | Curagen Corporation (US)  |  |
| source                     | Location/Qualifiers   |  |
|                            | 1..2565   |  |
| ORIGIN                     | /organism="Homo sapiens"  |  |
|                            | /mol_type="unassigned DNA"  |  |
|                            | /db_xref="taxon:9606"   |  |
| Query Match                | 35.9%;  | Score 2551; DB 6; Length 2565;                                     |
| Best Local Similarity      | 99.8%;  | Pred. No. 0;   |
| Matches 2554; Conservative | 0; Mismatches 5; Indels 0; Gaps 0;  |  |
| QY                         | 3262  | CCTGAAGGGTTCTTGGAAACCTGACTACGAATCCATTTGCTCAACTGCAATAAATGTA 3321    |
| DB                         | 7   | CCTGAAGGGTTTGTGGAAACCTGACTTACGAATCAATTTGCTCAACTGCAATAAATGTA 66     |
| QY                         | 3322  | AGCTGGGTCCACCGGCTCAACCAACGGTCTAGTCTTCTACTATGTTTCTCACTGATCTTA 3381  |
| DB                         | 67  | AGCTGGGTCCACCGGCTCAACCAACGGTCTAGTCTTCTACTATGTTTCTCACTGATCTTA 126   |
| QY                         | 3382  | CAGCAGACTCTCGGCATGTGAGACCACTCTTGTGTTACATATGAGAGAGCATATATTTT 3441   |
| DB                         | 127   | CAGCAGACTCTCGGCATGTGAGACCACTCTTGTGTTACATATGAGAGAGCATATATTTT 186    |
| QY                         | 3442  | GATAATCTGGAAAAATACACTGATATATATTTAAAAATTTACTCCATCAACAGAAAAGGGA 3501 |
| DB                         | 187   | GATAATCTGGAAAAATACACTGATATATATTTAAAAATTTACTCCATCAACAGAAAAGGGA 246  |
| QY                         | 3502  | TTCTCTGATACCTATACCTGCCAGCTATACATCAAGACTGGAAGAGATGTCCCGAGAACT 3561  |
| DB                         | 247   | TTCTCTGATACCTATACCTGCCAGCTATACATCAAGACTGGAAGAGATGTCCCGAGAACT 306   |
| QY                         | 3562  | TACCAATATCAACTTTTAAAAACCTTCTCTACCTGATCTCTTATATCATGGGAT 3621        |
| DB                         | 307   | TACCAATATCAACTTTTAAAAACCTTCTCTACCTGATCTCTTATATCATGGGAT 366         |
| QY                         | 3622  | CCCCAGTAAGCCAAATGTCATATATATTTAAAAATTTACTCCATCAACAGAACCAAT 3681     |
| DB                         | 367   | CCCCAGTAAGCCAAATGTCATATATATTTAAAAATTTACTCCATCAACAGAACCAAT 426      |
| QY                         | 3682  | GAATAATTTCTTCTCAATTTCTGATATATACATAATTTGGAAGAGCTTTCCACCATTT 3741    |
| DB                         | 427   | GAATAATTTCTTCTCAATTTCTGATATATACATAATTTGGAAGAGCTTTCCACCATTT 486     |
| QY                         | 3742  | ACATTATATAGCTTTTCTGCGGCAAGAACTAGAAAAGGACTGTGTCCTTCAGATTT 3801      |
| DB                         | 487   | ACATTATATAGCTTTTCTGCGGCAAGAACTAGAAAAGGACTGTGTCCTTCAGATTT 546       |
| QY                         | 3802  | CTTTTCTTTTACACAGATGAGTCAGTCGGTGTAGCACTCCACAAAATTTGACTTTAATC 3861   |
| DB                         | 547   | CTTTTCTTTTACACAGATGAGTCAGTCGGTGTAGCACTCCACAAAATTTGACTTTAATC 606    |
| QY                         | 3862  | AACTGTACTTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGGTATT 3921  |
| DB                         | 607   | AACTGTACTTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGGTATT 666   |
| QY                         | 3922  | GTTTAAAGTATATAGTTTTTAAAAATCATGAACATGAATGACCTATATATATTAAGAAT 3981   |

```
QY 5062 ATACCAGATGAGTTACAAAATTTCAATTAACGTTCTCTCCCTCTCTCAACTAATGCA 5121
Db 1807 ATACCAGATGAGTTACAAAATTTCAATTAACGTTCTCTCCCTCTCTCAACTAATGCA 1866

QY 5122 AATATCCAAAGTATATCAAGCTCTGGTTTACCGAGAAGATGATCTACTGCTGTCCAGATT 5181
Db 1867 AATATCCAAAGTATATCAAGCTCTGGTTTACCGAGAAGATGATCTACTGCTGTCCAGATT 1926

QY 5182 CACAACCTCAGTATTATACAGAAAACCAACACATTGCTCAATTTGCAATGCTAGAGACTA 5241
Db 1927 CACAACCTCAGTATTATACAGAAAACCAACACATTGCTCAATTTGCAATGCTAGAGACTA 1986

QY 5242 AAGGTTGGACATACATACATATACAGTGTGTTAGCGAGTCAATAGTCTGTGGTCAGGTCCA 5301
Db 1987 AAGGTTGGACATACATACATATACAGTGTGTTAGCGAGTCAATAGTCTGTGGTCAGGTCCA 2046

QY 5302 AAGGTTCCGATGAGAATAACCATGGATATCAAAAGCTCCAGCAGCAACAAAACCAACCA 5361
Db 2047 AAGGTTCCGATGAGAATAACCATGGATATCAAAAGCTCCAGCAGCAACAAAACCAACCA 2106

QY 5362 ACCCTTATTATGATCCACAGAAAACCTGTTGTCATCTTCAACAAATTAACAATCAGA 5421
Db 2107 ACCCTTATTATGATCCACAGAAAACCTGTTGTCATCTTCAACAAATTAACAATCAGA 2166

QY 5422 ATGCCAATATGTTACTACAGTGTATGATCATGACCAATTAATAAATGTACAAGTCTGCG 5481
Db 2167 ATGCCAATATGTTACTACAGTGTATGATCATGACCAATTAATAAATGTACAAGTCTGCG 2226

QY 5482 ACAGAAACAGGAGCTCAGCATGATGAAATGTAAACAAAGTGGTATGATGATCATATTTAAT 5541
Db 2227 ACAGAAACAGGAGCTCAGCATGATGAAATGTAAACAAAGTGGTATGATGATCATATTTAAT 2286

QY 5542 AAGCAAGGCCATATTTTACAAATGAAGGCTTTCTTAAACCTCCATGTACAGAGAAAG 5601
Db 2287 AAGCAAGGCCATATTTTACAAATGAAGGCTTTCTTAAACCTCCATGTACAGAGAAAG 2346

QY 5602 ACAAGCTTTAGTGGCAATGAAGAAATCTACATAGTGTGCTGATTAATGCATCATGATT 5661
Db 2347 ACAAGCTTTAGTGGCAATGAAGAAATCTACATAGTGTGCTGATTAATGCATCATGATT 2406

QY 5662 CTGGCAATGAACAAAAATTTGCAATGGACCACTGAAACCAAAAAAGCAATATTTATTT 5721
Db 2407 CTGGCAATGAACAAAAATTTGCAATGGACCACTGAAACCAAAAAAGCAATATTTATTT 2466

QY 5722 AATTTAGAGCTACAAATATTTATGGACAAATTTACTGACTGATTAATTTCTGACCTGTT 5781
Db 2467 AATTTAGAGCTACAAATATTTATGGACAAATTTACTGACTGATTAATTTCTGACCTGTT 2526

QY 5782 AAGACTTTAGGGGAAGGACTTTTCAGAAAGNACCGTAGAG 5820
Db 2527 AAGACTTTAGGGGAAGGACTTTTCAGAAAGNACCGTCGAG 2565

RESULT 11
LOCUS CQ726350
DEFINITION Sequence 12284 from Patent WO02068579.
ACCESSION CQ726350
VERSION CQ726350.1 GI:42289149
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Venter, C. J., Adams, M. C., Li, P. W. and Myers, E. W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 12284 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
Location/Qualifiers
```

```
source 1. .3149
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 31.7%; Score 2251; DB 6; Length 3149;
Best Local Similarity 80.5%; Pred. No. 0;
Matches 3020; Conservative 0; Mismatches 5; Indels 726; Gaps 4;

QY 2784 GAATAGATCATCATTAATAAACTATTAAATGTCATCTGAAACATCATTTGGAGATTATCAGATTT 2843
Db 1 GAATAGATCATCATTAATAAACTATTAAATGTCATCTGAAACATCATTTGGAGATTATCAGATTT 60

QY 2844 GAATATTAATGTTGAATACAGTCTTATGTATACAGCTAGCACCAAGATTTGGTGATGGAA 2903
Db 61 GGATATTAATGTTGAATACAGTCTTATGTATACAGCTAGCACCAAGATTTGGTGATGGAA 120

QY 2904 AACAGAGCAATATATCATTTAGCTTTTCAACACAGAGGAGGACCAAGCGATCCTCCCAA 2963
Db 121 AACAGAGCAATATATCATTTAGCTTTTCAACACAGAGGAGGACCAAGCGATCCTCCCAA 180

QY 2964 AGATGTTTATTATGCAAACTCAGTTCTTCAATCAATTAATTTCTTCTGACACCTCTTC 3023
Db 181 AGATGTTTATTATGCAAACTCAGTTCTTCAATCAATTAATTTCTTCTGACACCTCTTC 240

QY 3024 AAAACCTAATGGATATACATATTAATTTACTCTGTTTATACAGAAATACCTTCAGGTACTTT 3083
Db 241 AAAACCTAATGGATATACATATTAATTTACTCTGTTTATACAGAAATACCTTCAGGTACTTT 300

QY 3084 TATGCAAGATTTTACACTCCATGAACCTAATCAATGACTTTGACAATATGACTGTATCCAC 3143
Db 301 TATGCAAGATTTTACACTCCATGAAGTAACTAATGACTTTGACATATGACTGTATCCAC 360

QY 3144 AATTATAGATAAAGTACAAATATTCAGCTATATACATTTTGGTTAAACAGAGTACTTC 3203
Db 361 AATTATAGATAAAGTACAAATATTCAGCTATATACATTTTGGTTAAACAGAGTACTTC 420

QY 3204 AGTTGGAATGGGAATAAAGCAGTGACATCAATGGAAGTATACACAGATCAAGACATACC 3263
Db 421 AGTTGGAATGGGAATAAAGCAGTGACATCAATGGAAGTATACACAGATCAAGACATACC 476

QY 3264 TGAAGGGTTGTTGGAAACCTGACTTACGAATCCATTTCTGCTCAACTGCAATTAATGAAG 3323
Db 477 ----- 476

QY 3324 CTGGGTCCACCGGCTCAACAAAGGTCTAGTCTTCTACTATGTTTCACTGATCTTACA 3383
Db 477 ----- 476

QY 3384 GCAGACTCTCGCATGTGAGACCCTCTTGTGTATATGAGAGAAGCATATATTTTGA 3443
Db 477 ----- 476

QY 3444 TAATCTGGAAAAATAACACTGATTATATATAATAAATAATTAATCTCCATCAACAGAAAAAGGATT 3503
Db 477 ----- 476

QY 3504 CTCTGATACCTATATCTGCCAGCTATACATCAAGACTGAAGAAGATGTCCAGAAACTTC 3563
Db 477 -----TCCAGAAAACTTC 489

QY 3564 ACCAATAATCAACACTTTTAAAAACCTTTCTCTACCTCAGTTCTCTTATCATGGGATCC 3623
Db 490 ACCAATAATCAACACTTTTAAAAACCTTTCTCTACCTCAGTTCTCTTATCATGGGATCC 549

QY 3624 CCAGTAAAGCCAAATGGTGCAATATAAGTTATGATTTAACTTTTAAAGGACCAATGA 3683
Db 550 CCAGTAAAGCCAAATGGTGCAATATAAGTTATGATTTAACTTTTAAAGGACCAATGA 609

QY 3684 AAATTATTTCTTCAATTTCTTCTGATTAATTAATAATTTGGAAGAGCTTTCCACATTTAC 3743
Db 610 AAATTATTTCTTCAATTTCTTCTGATTAATTAATAATTTGGAAGAGCTTTCCACATTTAC 669
```

QY 3744 ATTATATAGCTTTTCTGCTCCGCAAGAACTAGAAAAGGACTTGGTCTCTTCCAGTATCT 3803  
DB 670 ATTATATAGCTTTTCTGCTCCGCAAGAACTAGAAAAGGACTTGGTCTCTTCCAGTATCT 729  
QY 3804 TTTCTTTTACACAGATGAGTCAGTCGCGTTAGCACCTCCACAAAATTTGACTTTTAATCAA 3863  
DB 730 TTTCTTTTACACAGATGAGTCAGTCGCGTTAGCACCTCCACAAAATTTGACTTTTAATCAA 789  
QY 3864 CTGTACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCCTCTTCCAGGTGGTATGT 3923  
DB 790 CTGTACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCCTCTTCCAGGTGGTATGT 849  
QY 3924 TAAAGTATATAGTTTAAATTCATGAACATGAACATGAACATGAACATGAACATGAACAT 3983  
DB 850 TAAAGTATATAGTTTAAATTCATGAACATGAACATGAACATGAACATGAACATGAACAT 909  
QY 3984 ATCAGGATTTAAACTGAAGCCAACTTGTGGCTGGAACCTGGAACCTGGAACCTGGAACCT 4043  
DB 910 ATCAGGATTTAAACTGAAGCCAACTTGTGGCTGGAACCTGGAACCTGGAACCTGGAACCT 969  
QY 4044 CCGTGATCTGCGTTCCAAAGATGGAAATGGCAATCAATTTAGTAATGTAGTAAAT 4103  
DB 970 CCGTGATCTGCGTTCCAAAGATGGAAATGGCAATCAATTTAGTAATGTAGTAAAT 1029  
QY 4104 CACAAACCAAGATCAGTTCAGATGTCGTCGAGAAATGCAATGCAATGCAATGCAATGCAAT 4163  
DB 1030 CACAAACCAAGATCAGTTCAGATGTCGTCGAGAAATGCAATGCAATGCAATGCAATGCAAT 1089  
QY 4164 GCAGTCAGTTTGTAGTGAATGGGATCCACCCAAAGAGGCAATGGAAATGAATGAATGAAT 4223  
DB 1090 GCAGTCAGTTTGTAGTGAATGGGATCCACCCAAAGAGGCAATGGAAATGAATGAATGAAT 1149  
QY 4224 TATGGTAAACAGTTGAAAGGAATCTACAAAGTTTCTCCCAAGATCACATGTACACTTT 4283  
DB 1150 TATGGTAAACAGTTGAAAGGAATCTACAAAGTTTCTCCCAAGATCACATGTACACTTT 1209  
QY 4284 CATAAAGCTTTGCGCAATACCTCATATGTCCTTTAAAGTAAGAGTTCAACCTCAGCTGG 4343  
DB 1210 CATAAAGCTTTGCGCAATACCTCATATGTCCTTTAAAGTAAGAGTTCAACCTCAGCTGG 1269  
QY 4344 TGAAGGTGAAGACATGCCATGTCAGCACATCTACCTGAAACAGATTTCCAGTGTCC 4403  
DB 1270 TGAAGGTGAAGACATGCCATGTCAGCACATCTACCTGAAACAGATTTCCAGTGTCC 1316  
QY 4404 CACAAATATTGCTTTCTGATGTTCACTGATGTCAGTCAACTAGTGCAACATTGACATGGATAAGACC 4463  
DB 1317 ----- 1316  
QY 4464 TGACACTATCTTTGGCTACTTTCAAAATTAACAAATTAACCACTTCCAGTGTCCAAA 4523  
DB 1317 ----- 1316  
QY 4524 ATGCAAGATGGGAATCCGAAGATGTTGTAATATCAAAAATTCATACCTCTATGA 4583  
DB 1317 ----- 1316  
QY 4584 AGCTCACTTAACCTGAAGAGACAGTATATGATTAAGAAATTTAGATGGTATAGATTCCA 4643  
DB 1317 ----- 1316  
QY 4644 AGTGGCTGCAGCACCACATGCTGGCTATGGCAATGCTTCAAACTGGATTTCTACAAAAAC 4703  
DB 1317 ----- 1316  
QY 4704 TCTGCTGGCCCTCCAGATGGTCTCTGTAATGTTTATGATGATGATGATGATGATGATGATGAT 4763  
DB 1317 -----CTCCAGATGGTCTCTGTAATGTTTATGATGATGATGATGATGATGATGATGAT 1365  
QY 4764 TAGCATCAGCATAGCTGGAGTCAACTGCTGTCATTTACTGGACCAACATGTTTATCTGAT 4823  
DB 1366 TAGCATCAGCATAGCTGGAGTCAACTGCTGTCATTTACTGGACCAACATGTTTATCTGAT 1425

QY 4824 TGATGTCAAAATCGGTAGATAATGATGAATTTAATATATATCTTCTCATCAAGTCAAAATGAAGA 4883  
DB 1426 TGATGTCAAAATCGGTAGATAATGATGAATTTAATATATATCTTCTCATCAAGTCAAAATGAAGA 1485  
QY 4884 AATAAAAAACCATAGAAATTAAGATTTAGAAATATTTACAAAGGTATTTCTGTAGTGTATC 4943  
DB 1486 AATAAAAAACCATAGAAATTAAGATTTAGAAATATTTACAAAGGTATTTCTGTAGTGTATC 1545  
QY 4944 TGCAATTTACTGGGAACATTTAGTGTCTCATATGTAGAGGGAAGTCAAGTGTCTGAAATGAT 5003  
DB 1546 TGCAATTTACTGGGAACATTTAGTGTCTCATATGTAGAGGGAAGTCAAGTGTCTGAAATGAT 1605  
QY 5004 TGTTACTACTTTAGATTCAGCCCCCAAGGACCCCACTAACACATGACATTTTCAGAGAT 5063  
DB 1606 TGTTACTACTTTAGATTC----- 1623  
QY 5064 ACCAGATGAAGTTTACAAAATTTCAAATTAAGTTTCTTCTCTCTCTCAACCTAATGGAAA 5123  
DB 1624 ----- 1623  
QY 5124 TATCCAGTATATCAAGCTCTGGTTTACCGAGAAGATGATCTCTACTGTCCAGATTTCA 5183  
DB 1624 -----AGTATATCAAGCTCTGGTTTACCGAGAAGATGATCTCTACTGTCTCCAGATTTCA 1677  
QY 5184 CAACCTCAGTATTTATACAGAAAACCAACACATTCGTCTGCAATGCTAGAGAGACTAAA 5243  
DB 1678 CAACCTCAGTATTTATACAGAAAACCAACACATTCGTCTGCAATGCTAGAGAGACTAAA 1737  
QY 5244 AGGTGACATACATACATATCAATGCTGTTTACCGAGTCAATAGTGTCTGAGTGTCTCAAA 5303  
DB 1738 AGGTGACATACATACATATCAATGCTGTTTACCGAGTCAATAGTGTCTGAGTGTCTCNA 1797  
QY 5304 GGTTCGATGAGAATAACCATGATATCAAGCTCCAGACGACCAAAACCAAAACCAAC 5363  
DB 1798 GGTTCGATGAGAATAACCATGATATCAAGCTCCAGACGACCAAAACCAAAACCAAC 1857  
QY 5364 CCTATTTATGATGCCACAGGAAAACCTGTTGTGACTTCAACAAATTTCAATTCAGAT 5423  
DB 1858 CCTATTTATGATGCCACAGGAAAACCTGTTGTGACTTCAACAAATTTCAATTCAGAT 1917  
QY 5424 GCCAATATGTTACTACAGTGTATCATGACCAATATAAATAATGTAACAGTGTCTCGAC 5483  
DB 1918 GCCAATATGTTACTACAGTGTATCATGACCAATATAAATAATGTAACAGTGTCTGTGAC 1977  
QY 5484 AGAAAACAGGAGCTCAGCATGATGGAATGTAAACAAAGTGGTATGATGATGATGATGATGAT 5543  
DB 1978 AGAAAACAGGAGCTCAGCATGATGGAATGTAAACAAAGTGGTATGATGATGATGATGAT 2037  
QY 5544 AGCAAGCCATATTTTACAAATGAAGCTTTCTTAACCTCCATGTACAGAGGAAAGAC 5603  
DB 2038 AGCAAGCCATATTTTACAAATGAAGCTTTCTTAACCTCCATGTACAGAGGAAAGAC 2097  
QY 5604 AAGTTTGTAGTGGCAATGAAGAAATCTACATCATAGTGTGATGATGATGATGATGATGATGAT 5663  
DB 2098 AAGTTTGTAGTGGCAATGAAGAAATCTACATCATAGTGTGATGATGATGATGATGATGAT 2157  
QY 5664 TGGCAATGAAGCAAAAATTTGCAATGGACCACTGAAACCAAAAGCAAAATCTTTATTA 5723  
DB 2158 TGGCAATGAAGCAAAAATTTGCAATGGACCACTGAAACCAAAAGCAAAATCTTTATTA 2217  
QY 5724 ATTTAGAGCTCAAAATTTATGGGCAATTTTACTGATGATGATGATGATGATGATGATGAT 5783  
DB 2218 ATTTAGAGCTCAAAATTTATGGGCAATTTTACTGATGATGATGATGATGATGATGATGAT 2277  
QY 5784 GACTTTAGGGGAAGGACTTTTACAGAAAGACCTGAGATGATGATGATGATGATGATGATGAT 5843  
DB 2278 GACTTTAGGGGAAGGACTTTTACAGAAAGACCTGAGATGATGATGATGATGATGATGATGAT 2337  
QY 5844 TATCTTTTCAATAATTTCTCTTGGGAACAGCTATTTTGGCAATTTTGGCAATTTTGGCA 5903  
DB 2338 TATCTTTTCAATAATTTCTCTTGGGAACAGCTATTTTGGCAATTTTGGCAATTTTGGCA 2397  
QY 5904 GCAGAAAGAGGTGGGACATATCTCTCTCAGGATGAGAAATTTATTTGACACTAAATTTGAA 5963

|                            |  |   |                            |
|----------------------------|--|---|----------------------------|
| Db                         | 2398   | GCAGAAAGAGGTGGACATATCTCTCCTCAGGTCGAGAAATATTGACACTAAATTGAA       | 2457                       |
| Qy                         | 5964   | GCTGGATCAGCTCATCAGTGGCAGACCTGGAACTGAAGGACGAGAGATTAAACGGCC       | 6023                       |
| Db                         | 2458   | GCTGGATCAGCTCATCAGTGGCAGACCTGGAACTGAAGGACGAGAGATTAAACGGCC       | 2517                       |
| Qy                         | 6024   | AATAGCAGAAATCCTTCTCTGCAACATGTTGAGAGCTTTGCACAAACAACCTTAA         | 6083                       |
| Db                         | 2518   | AATAGCAGAAATCCTTCTCTGCAACATGTTGAGAGCTTTGCACAAACAACCTTAA         | 2577                       |
| Qy                         | 6084   | GTTCCTCAAGAGAAATTTTCGGAATTTACCAAAATTTCTTCAGGATCTTTCTTCAACTGATGC | 6143                       |
| Db                         | 2578   | GTTCCTCAAGAGAAATTTTCGGAATTTACCAAAATTTCTTCAGGATCTTTCTTCAACTGATGC | 2637                       |
| Qy                         | 6144   | TGATCTGCTTGGATAGAGCAAAAAACCGTTTCCCAACATAAACCATAAATAA            | 6203                       |
| Db                         | 2638   | TGATCTGCTTGGATAGAGCAAAAAACCGTTTCCCAACATAAACCATAAATAA            | 2694                       |
| Qy                         | 6204   | TACAGAGTAAGCTGATAGCTGACGTAGTGTTCAGGTTCCGATTATTAATGCCAG          | 6263                       |
| Db                         | 2695   | TACAGAGTAAGCTGATAGCTGACGTAGTGTTCAGGTTCCGATTATTAATGCCAG          | 2754                       |
| Qy                         | 6264   | CTATATTTCTGGTTATTTATGTCCTAAATGAATTTATTGCTACTCAAGGTCCTACTACGAG   | 6323                       |
| Db                         | 2755   | CTATATTTCTGGTTATTTATGTCCTAAATGAATTTATTGCTACTCAAGGTCCTACTACGAG   | 2814                       |
| Qy                         | 6324   | AACAGTTGGAGATTTTGGAGAAATGTTGGGAAACCGAGGCAAAACAATTAGTAATGCT      | 6383                       |
| Db                         | 2815   | AACAGTTGGAGATTTTGGAGAAATGTTGGGAAACCGAGGCAAAACAATTAGTAATGCT      | 2874                       |
| Qy                         | 6384   | AACAGAGTTTGAAGAGGACGATCAGATGCCATCAGTATTGGCCAGAGGACAA            | 6443                       |
| Db                         | 2875   | AACAGAGTTTGAAGAGGACGATCAGATGCCATCAGTATTGGCCAGAGGACAA            | 2934                       |
| Qy                         | 6444   | GCAGGTTACTGCTTTGGAGATATAGTGATTACAAAGCTAAATGGAGGATTTCAATAGA      | 6503                       |
| Db                         | 2935   | GCAGGTTACTGCTTTGGAGATATAGTGATTACAAAGCTAAATGGAGGATTTCAATAGA      | 2994                       |
| Qy                         | 6504   | TTGGACTATCAGGATCTGAAATTCGAAAGG                                  | 6534                       |
| Db                         | 2995   | TTGGACTATCAGGATCTGAAATTCGAAAGG                                  | 3025                       |
| RESULT 12                  |  |   |                            |
| LOCUS                      | AR073854   | 2692 bp   | DNA linear PAT 28-AUG-2000 |
| DEFINITION                 | Sequence 14 from patent US 5952212.                    |   |                            |
| ACCESSION                  | AR073854   |   |                            |
| VERSION                    | AR073854.1 GI:10000614                                 |   |                            |
| KEYWORDS                   | Unknown.   |   |                            |
| SOURCE                     | Unknown.   |   |                            |
| ORGANISM                   | Unclassified.  |   |                            |
| REFERENCE                  | 1 (bases 1 to 2692)                                    |   |                            |
| AUTHORS                    | Moller,N.Peter.Hundahl., Moller,K.Bach. and Ullrich,A. |   |                            |
| TITLE                      | Protein tyrosine phosphatase                           |   |                            |
| JOURNAL                    | Patent: US 5952212-A 14 14-SEP-1999;                   |   |                            |
| FEATURES                   | Location/Qualifiers                                    |   |                            |
| source                     | 1..2692  |   |                            |
|                            | /organism="unknown"                                    |   |                            |
|                            | /mol_type="unassigned DNA"                             |   |                            |
| ORIGIN                     |  |   |                            |
| Query Match                | 15.7%; Score 1118.8; DB 6; Length 2692;                |   |                            |
| Best Local Similarity      | 93.7%; Pred. No. 1.2e-216;                             |   |                            |
| Matches 1218; Conservative | 0; Mismatches 7; Indels 75; Gaps 2;                    |   |                            |
| Qy                         | 5881   | GCATTTGCAAGATTCGACAGAGAGCGAAAGAGGTGGCACAATACATCTCTCTCAGGATGCA   | 5940                       |
| Db                         | 247  | GCTTGTGTTAGAAATTCGACAGAGAGCGAAAGAGGTGGCACAATACATCTCTCTCAGGATGCA | 306                        |
| Qy                         | 5941   | GAAATTTATGACACTAAATTTGAAGCTGGATCAGCTCATCAGTGGCAGACCTGGAACTG     | 6000                       |

|    |      |   |      |
|----|------|---|------|
| Db | 307  | GAAATTTATGACACTAAATTTGAAGCTGGATCAGCTCATCAGTGGCAGACCTGGAACTG     | 366  |
| Qy | 6001 | AAGGACGAGAGATTAAACGGC-----                                      | 6020 |
| Db | 367  | AAGGACGAGAGATTAAACGGACTACTCTCTCAITTTTCTTTAGACGCAAGGAGATTTTGTGTC | 426  |
| Qy | 6021 | -----GCCAAATAAGCAAGAAATCCTTCTCTGCAA                             | 6048 |
| Db | 427  | ATCCAGTTACTTAGTTATAGAAAATCCATCAAGCCCAATAGCAAGAAATCCTTCTCTGCAA   | 486  |
| Qy | 6049 | CATGTTTGAAGAGCTTTGACACAAACAACCTTAAAGTTTCAAGAAAGAAATTTTCGAAATTA  | 6108 |
| Db | 487  | CATGTTTGAAGAGCTTTGACACAAACAACCTTAAAGTTTCAAGAAAGAAATTTTCGAAATTA  | 546  |
| Qy | 6109 | CCAAAATTTCTTCAGGATCTTTCTTCAACTGATGCTGCTGCTTGGAAATAGACAAA        | 6168 |
| Db | 547  | CCAAAATTTCTTCAGGATCTTTCTTCAACTGATGCTGCTGCTTGGAAATAGACAAA        | 606  |
| Qy | 6169 | RACCGTTTCCCAAAACATATAATATAATAACAGAGTAAAGCTGATAGCTGAC            | 6228 |
| Db | 607  | RACCGTTTCCCAAAACATATAATATAATAACAGAGTAAAGCTGATAGCTGAC            | 663  |
| Qy | 6229 | GCTAGTGTTCAGAGTTCCGATTTATTAATGCGAGCTATATTTCTGGTTATTTATGTC       | 6288 |
| Db | 664  | GCTAGTGTTCAGAGTTCCGATTTATTAATGCGAGCTATATTTCTGGTTATTTATGTC       | 723  |
| Qy | 6289 | AATGAATTTATGCTACTCAAGTCCACTACAGAGAACAGTTGGAGATTTTGGAGATG        | 6348 |
| Db | 724  | AATGAATTTATGCTACTCAAGTCCACTACAGAGAACAGTTGGAGATTTTGGAGATG        | 783  |
| Qy | 6349 | GTGTCGGAAACCCAGGCGCAAAACATTAGTAATGCTTAAACACAGTGTTTGAAAAAGACGG   | 6408 |
| Db | 784  | GTGTCGGAAACCCAGGCGCAAAACATTAGTAATGCTTAAACACAGTGTTTGAAAAAGACGG   | 843  |
| Qy | 6409 | ATCAGATGCCATCAGTATTTGGCCAGAGGACAAACAAGCCAGTTACTGTCTTTGGAGATATA  | 903  |
| Db | 844  | ATCAGATGCCATCAGTATTTGGCCAGAGGACAAACAAGCCAGTTACTGTCTTTGGAGATATA  | 963  |
| Qy | 6469 | GTGATTAACAAGCTAATGGAGATGTTCAATAGATTTGAGCTATCAGGGATCTGAAATTT     | 6528 |
| Db | 904  | GTGATTAACAAGCTAATGGAGATGTTCAATAGATTTGAGCTATCAGGGATCTGAAATTT     | 963  |
| Qy | 6529 | GAAAGCATGGGATTTGCATGACTGTTCGACAGTGTAACTTTTACTGCTGCGGACAGCAT     | 6588 |
| Db | 964  | GAAAGCATGGGATTTGCATGACTGTTCGACAGTGTAACTTTTACTGCTGCGGACAGCAT     | 1023 |
| Qy | 6589 | GGGGTTCTTGAGAAACAGCGCCCTCTAAATTCATCTTTGTGAAGTTGGTTGAGCAAGCAGG   | 6648 |
| Db | 1024 | GGGGTTCTTGAGAAACAGCGCCCTCTAAATTCATCTTTGTGAAGTTGGTTGAGCAAGCAGG   | 1083 |
| Qy | 6649 | GCATATGACACACACTATGATTTGTTCTGTCAGTGTGGAGTTGGAAAGAACTGGAGTT      | 6708 |
| Db | 1084 | GCATATGACACACACTATGATTTGTTCTGTCAGTGTGGAGTTGGAAAGAACTGGAGTT      | 1143 |
| Qy | 6709 | TTTATTTGCTCTGACCACTTTTACACAAACATATAAATGACCATGATTTTGTGGATATAT    | 6768 |
| Db | 1144 | TTTATTTGCTCTGACCACTTTTACACAAACATATAAATGACCATGATTTTGTGGATATAT    | 1203 |
| Qy | 6769 | GGACTAGTAGCTGAACTGAGAAAGTGAAGAAATGTGCATGGTGCAGATCTGACACAGTAT    | 6828 |
| Db | 1204 | GGACTAGTAGCTGAACTGAGAAAGTGAAGAAATGTGCATGGTGCAGATCTGACACAGTAT    | 1263 |
| Qy | 6829 | ATCTTTTACACAGTGCATTTCTGATCTCTTATCAATAGGGAGTAAATCAGCCCATC        | 6888 |
| Db | 1264 | ATCTTTTACACAGTGCATTTCTGATCTCTTATCAATAGGGAGTAAATCAGCCCATC        | 1323 |
| Qy | 6889 | TGTTTTGTAACTATTTCAGCACTTCAGAGATGGACTCTTTTCGACGCCATGGAAGTGAT     | 6948 |
| Db | 1324 | TGTTTTGTAACTATTTCAGCACTTCAGAGATGGACTCTTTTCGACGCCATGGAAGTGAT     | 1383 |
| Qy | 6949 | GTTCAGCTTTGAATGGGAGAAACCACTATGTAAATTTTCAGACCAAGAGATCAATTTGA     | 7008 |





| Matches 1087; Conservative 0; Mismatches 4; Indels 3; Gaps 1; |      |  |      |  |  |  |  |  |  |
|---|------|--|------|--|--|--|--|--|--|
| QY  | 6015 | AAAGCGGCAATAAGCAAGAAATCCCTCTCGCAACATGTTGAAGAGCTTTGCACAAACA     | 6074 |  |  |  |  |  |  |
| Db  | 70   | AATGAGGCCAATAAGCAAGAAATCCCTCTCGCAACATGTTGAAGAGCTTTGCACAAACA    | 129  |  |  |  |  |  |  |
| QY  | 6075 | CAACCTAAAGTTTCAAGAGAAATTTTCGGAAATTTACCAAAATTTCTTCAGGATCTTTCTTC | 6134 |  |  |  |  |  |  |
| Db  | 130  | CAACCTAAAGTTTCAAGAGAAATTTTCGGAAATTTACCAAAATTTCTTCAGGATCTTTCTTC | 189  |  |  |  |  |  |  |
| QY  | 6135 | AACTGATGCTGATCTCCCTTGGAAATAGAGCAAAAACCGTTTCCCAAAACATAAAACCATP  | 6194 |  |  |  |  |  |  |
| Db  | 190  | AACTGATGCTGATCTCCCTTGGAAATAGAGCAAAAACCGTTTCCCAAAACATAAAACCATP  | 248  |  |  |  |  |  |  |
| QY  | 6195 | TAATATAATACAGAGTAAGCTGATGCTGACGTAGTGTTCAGAGTTCCGATTTATAT       | 6254 |  |  |  |  |  |  |
| Db  | 249  | --ATAATAATAACAGAGTAAGCTGATGCTGACGTAGTGTTCAGAGTTCCGATTTATAT     | 306  |  |  |  |  |  |  |
| QY  | 6255 | TAATGTCAGCTATATTTCTGGTTTATTTATGTCCTCAATGAATTTATTTGCTCAAGGTC    | 6314 |  |  |  |  |  |  |
| Db  | 307  | TAATGTCAGCTATATTTCTGGTTTATTTATGTCCTCAATGAATTTATTTGCTCAAGGTC    | 366  |  |  |  |  |  |  |
| QY  | 6315 | ACTACAGGAACAGTTGGAGATTTTGGAGAAATGGTGGGAAACAGGCGCAAAAACATP      | 6374 |  |  |  |  |  |  |
| Db  | 367  | ACTACAGGAACAGTTGGAGATTTTGGAGAAATGGTGGGAAACAGGCGCAAAAACATP      | 426  |  |  |  |  |  |  |
| QY  | 6375 | AGTAATGCTAACACAGTGTTTGAAAAGGACCGATCAGATGCCATCAGTATTTGGCCAGA    | 6434 |  |  |  |  |  |  |
| Db  | 427  | AGTAATGCTAACACAGTGTTTGAAAAGGACCGATCAGATGCCATCAGTATTTGGCCAGA    | 486  |  |  |  |  |  |  |
| QY  | 6435 | GGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGATTAACAAGCTAATCGAGGATGT   | 6494 |  |  |  |  |  |  |
| Db  | 487  | GGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGATTAACAAGCTAATCGAGGATGT   | 546  |  |  |  |  |  |  |
| QY  | 6495 | TCAAATAGATTGGAGCTATCAGGGATCTGAAAATTTGAAAGGCAATGGGATTTGCATGCT   | 6554 |  |  |  |  |  |  |
| Db  | 547  | TCAAATAGATTGGAGCTATCAGGGATCTGAAAATTTGAAAGGCAATGGGATTTGCATGCT   | 606  |  |  |  |  |  |  |
| QY  | 6555 | TCGACAGTGTAACTTTACTGCTGCGCAGAGCATGATGACCAACCTATGATTTCT         | 6614 |  |  |  |  |  |  |
| Db  | 607  | TCGACAGTGTAACTTTACTGCTGCGCAGAGCATGATGACCAACCTATGATTTCT         | 666  |  |  |  |  |  |  |
| QY  | 6615 | AATTCACTTTGTAAGTTGGTTCGACAGAGGCGACATGACCAACCTATGATTTCT         | 6674 |  |  |  |  |  |  |
| Db  | 667  | AATTCACTTTGTAAGTTGGTTCGACAGAGGCGACATGACCAACCTATGATTTCT         | 726  |  |  |  |  |  |  |
| QY  | 6675 | TCACGTCAGTGTGGAGTTGGAGAACTGGAGTTTATTTGCTCTGGACCAATTAACACA      | 6734 |  |  |  |  |  |  |
| Db  | 727  | TCACGTCAGTGTGGAGTTGGAGAACTGGAGTTTATTTGCTCTGGACCAATTAACACA      | 786  |  |  |  |  |  |  |
| QY  | 6735 | ACATAAATGACCATGATTTTGTGGATATATATGACCTAGTGTGAATCTGAGAAAGTGA     | 6794 |  |  |  |  |  |  |
| Db  | 787  | ACATAAATGACCATGATTTTGTGGATATATATGACCTAGTGTGAATCTGAGAAAGTGA     | 846  |  |  |  |  |  |  |
| QY  | 6795 | AGAATGTGATGCTGCAGATCTGGCAGAGTATATCTTTTACACAGTGCATCTGGA         | 6854 |  |  |  |  |  |  |
| Db  | 847  | AGAATGTGATGCTGCAGATCTGGCAGAGTATATCTTTTACACAGTGCATCTGGA         | 906  |  |  |  |  |  |  |
| QY  | 6855 | TCTCTTATCAAAATAAGGGAAGTAATCAGGCCATCTGTTTGTAACTATTTCAGCACTTCA   | 6914 |  |  |  |  |  |  |
| Db  | 907  | TCTCTTATCAAAATAAGGGAAGTAATCAGGCCATCTGTTTGTAACTATTTCAGCACTTCA   | 966  |  |  |  |  |  |  |
| QY  | 6915 | GAAAGTGGACTCTTTGGACGCCATGGAAGTGATGTTGAGCTTGAATGGGAGAAACCAAC    | 6974 |  |  |  |  |  |  |
| Db  | 967  | GAAAGTGGACTCTTTGGACGCCATGGAAGTGATGTTGAGCTTGAATGGGAGAAACCAAC    | 1026 |  |  |  |  |  |  |
| QY  | 6975 | TATGTAATAATTTCAGACCAAAAGATACAAATGGAACAGATTTTAAATCCAGGGGCCAA    | 7034 |  |  |  |  |  |  |
| Db  | 1027 | TATGTAATAATTTCAGACCAAAAGATACAAATGGAACAGATTTTAAATCCAGGGGCCAA    | 1086 |  |  |  |  |  |  |
| QY  | 7035 | AGTTACCCCTCATTTCTTCGAAATGAAATGTGCAACCTTAAAGAAATATCTATGCTTCT    | 7094 |  |  |  |  |  |  |
| Db  | 1087 | AGTTACCCCTCATTTCTTCGAAATGAAATGTGCAACCTTAAAGAAATATCTATGCTTCT    | 1146 |  |  |  |  |  |  |

|                       |   |  |           |              |                 |  |  |  |  |
|-----------------------|---|--|-----------|--------------|-----------------|--|--|--|--|
| QY                    | 7095  | CTCACTGTGCGCTTT  | 7108      |              |                 |  |  |  |  |
| Db                    | 1147  | CTCACTGTGCGCTTT  | 1160      |              |                 |  |  |  |  |
| RESULT 15             |   |  |           |              |                 |  |  |  |  |
| I32037                |   |  |           |              |                 |  |  |  |  |
| LOCUS                 | I32037  | 2309 bp  | DNA       | linear       | PAT 06-FEB-1997 |  |  |  |  |
| DEFINITION            | Sequence 10 from patent US 5585233.           |  |           |              |                 |  |  |  |  |
| ACCESSION             | I32037  |  |           |              |                 |  |  |  |  |
| VERSION               | I32037.1                                      | GI:1822828   |           |              |                 |  |  |  |  |
| KEYWORDS              |   |  |           |              |                 |  |  |  |  |
| SOURCE                | Unknown.                                      |  |           |              |                 |  |  |  |  |
| ORGANISM              | Unclassified.                                 |  |           |              |                 |  |  |  |  |
| REFERENCE             | 1 (bases 1 to 2309)                           |  |           |              |                 |  |  |  |  |
| AUTHORS               | Moller,N.P.H., Moller,K.B. and Ullrich,A.     |  |           |              |                 |  |  |  |  |
| TITLE                 | PTP-S31: A novel protein tyrosine phosphatase |  |           |              |                 |  |  |  |  |
| JOURNAL               | Patent: US 5585233-A 10 17-DEC-1996;          |  |           |              |                 |  |  |  |  |
| FEATURES              | Location/Qualifiers                           |  |           |              |                 |  |  |  |  |
| source                | 1..2309                                       |  |           |              |                 |  |  |  |  |
|                       | /organism="unknown"                           |  |           |              |                 |  |  |  |  |
| ORIGIN                | /mol_type="unassigned DNA"                    |  |           |              |                 |  |  |  |  |
| Query Match           | 15.1%;  | Score 1071.6;  | DB 6;     | Length 2309; |                 |  |  |  |  |
| Best Local Similarity | 99.4%;  | Pred. No. 4.7e-207;  |           |              |                 |  |  |  |  |
| Matches 1087;         | Conservative 0;                               | Mismatches 4;  | Indels 3; | Gaps 1;      |                 |  |  |  |  |
| QY                    | 6015  | AAAGCGGCAATAAGCAAGAAATCCCTCTCGCAACATGTTGAAGAGCTTTGCACAAACA     | 6074      |              |                 |  |  |  |  |
| Db                    | 70  | AATGAGGCCAATAAGCAAGAAATCCCTCTCGCAACATGTTGAAGAGCTTTGCACAAACA    | 129       |              |                 |  |  |  |  |
| QY                    | 6075  | CAACCTAAAGTTTCAAGAGAAATTTTCGGAAATTTACCAAAATTTCTTCAGGATCTTTCTTC | 6134      |              |                 |  |  |  |  |
| Db                    | 130   | CAACCTAAAGTTTCAAGAGAAATTTTCGGAAATTTACCAAAATTTCTTCAGGATCTTTCTTC | 189       |              |                 |  |  |  |  |
| QY                    | 6135  | AACTGATGCTGATCTCCCTTGGAAATAGAGCAAAAACCGTTTCCCAAAACATAAAACCATP  | 6194      |              |                 |  |  |  |  |
| Db                    | 190   | AACTGATGCTGATCTCCCTTGGAAATAGAGCAAAAACCGTTTCCCAAAACATAAAACCATP  | 248       |              |                 |  |  |  |  |
| QY                    | 6195  | TAATATAATACAGAGTAAGCTGATGCTGACGTAGTGTTCAGAGTTCCGATTTATAT       | 6254      |              |                 |  |  |  |  |
| Db                    | 249   | --ATAATAATAACAGAGTAAGCTGATGCTGACGTAGTGTTCAGAGTTCCGATTTATAT     | 306       |              |                 |  |  |  |  |
| QY                    | 6255  | TAATGCCAGCTATATTTCTGGTTTATTTATGTCCTCAATGAATTTATTTGCTCAAGGTC    | 6314      |              |                 |  |  |  |  |
| Db                    | 307   | TAATGCCAGCTATATTTCTGGTTTATTTATGTCCTCAATGAATTTATTTGCTCAAGGTC    | 366       |              |                 |  |  |  |  |
| QY                    | 6315  | ACTACAGGAACAGTTGGAGATTTTGGAGAAATGGTGGGAAACAGGCGCAAAAACATP      | 6374      |              |                 |  |  |  |  |
| Db                    | 367   | ACTACAGGAACAGTTGGAGATTTTGGAGAAATGGTGGGAAACAGGCGCAAAAACATP      | 426       |              |                 |  |  |  |  |
| QY                    | 6375  | AGTAATGCTAACACAGTGTTTGAAAAGGACCGATCAGATGCCATCAGTATTTGGCCAGA    | 6434      |              |                 |  |  |  |  |
| Db                    | 427   | AGTAATGCTAACACAGTGTTTGAAAAGGACCGATCAGATGCCATCAGTATTTGGCCAGA    | 486       |              |                 |  |  |  |  |
| QY                    | 6435  | GGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGATTAACAAGCTAATCGAGGATGT   | 6494      |              |                 |  |  |  |  |
| Db                    | 487   | GGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGATTAACAAGCTAATCGAGGATGT   | 546       |              |                 |  |  |  |  |
| QY                    | 6495  | TCAAATAGATTGGAGCTATCAGGGATCTGAAAATTTGAAAGGCAATGGGATTTGCATGCT   | 6554      |              |                 |  |  |  |  |
| Db                    | 547   | TCAAATAGATTGGAGCTATCAGGGATCTGAAAATTTGAAAGGCAATGGGATTTGCATGCT   | 606       |              |                 |  |  |  |  |
| QY                    | 6555  | TCGACAGTGTAACTTTACTGCTGCGCAGAGCATGATGACCAACCTATGATTTCT         | 6614      |              |                 |  |  |  |  |
| Db                    | 607   | TCGACAGTGTAACTTTACTGCTGCGCAGAGCATGATGACCAACCTATGATTTCT         | 666       |              |                 |  |  |  |  |
| QY                    | 6615  | AATTCACTTTGTAAGTTGGTTCGACAGAGGCGACATGACCAACCTATGATTTCT         | 6674      |              |                 |  |  |  |  |
| Db                    | 667   | AATTCACTTTGTAAGTTGGTTCGACAGAGGCGACATGACCAACCTATGATTTCT         | 726       |              |                 |  |  |  |  |
| QY                    | 6675  | TCACGTCAGTGTGGAGTTGGAGAACTGGAGTTTATTTGCTCTGGACCAATTAACACA      | 6734      |              |                 |  |  |  |  |
| Db                    | 727   | TCACGTCAGTGTGGAGTTGGAGAACTGGAGTTTATTTGCTCTGGACCAATTAACACA      | 786       |              |                 |  |  |  |  |
| QY                    | 6735  | ACATAAATGACCATGATTTTGTGGATATATATGACCTAGTGTGAATCTGAGAAAGTGA     | 6794      |              |                 |  |  |  |  |
| Db                    | 787   | ACATAAATGACCATGATTTTGTGGATATATATGACCTAGTGTGAATCTGAGAAAGTGA     | 846       |              |                 |  |  |  |  |
| QY                    | 6795  | AGAATGTGATGCTGCAGATCTGGCAGAGTATATCTTTTACACAGTGCATCTGGA         | 6854      |              |                 |  |  |  |  |
| Db                    | 847   | AGAATGTGATGCTGCAGATCTGGCAGAGTATATCTTTTACACAGTGCATCTGGA         | 906       |              |                 |  |  |  |  |
| QY                    | 6855  | TCTCTTATCAAAATAAGGGAAGTAATCAGGCCATCTGTTTGTAACTATTTCAGCACTTCA   | 6914      |              |                 |  |  |  |  |
| Db                    | 907   | TCTCTTATCAAAATAAGGGAAGTAATCAGGCCATCTGTTTGTAACTATTTCAGCACTTCA   | 966       |              |                 |  |  |  |  |
| QY                    | 6915  | GAAAGTGGACTCTTTGGACGCCATGGAAGTGATGTTGAGCTTGAATGGGAGAAACCAAC    | 6974      |              |                 |  |  |  |  |
| Db                    | 967   | GAAAGTGGACTCTTTGGACGCCATGGAAGTGATGTTGAGCTTGAATGGGAGAAACCAAC    | 1026      |              |                 |  |  |  |  |
| QY                    | 6975  | TATGTAATAATTTCAGACCAAAAGATACAAATGGAACAGATTTTAAATCCAGGGGCCAA    | 7034      |              |                 |  |  |  |  |
| Db                    | 1027  | TATGTAATAATTTCAGACCAAAAGATACAAATGGAACAGATTTTAAATCCAGGGGCCAA    | 1086      |              |                 |  |  |  |  |
| QY                    | 7035  | AGTTACCCCTCATTTCTTCGAAATGAAATGTGCAACCTTAAAGAAATATCTATGCTTCT    | 7094      |              |                 |  |  |  |  |
| Db                    | 1087  | AGTTACCCCTCATTTCTTCGAAATGAAATGTGCAACCTTAAAGAAATATCTATGCTTCT    | 1146      |              |                 |  |  |  |  |



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2005, 19:01:59 ; Search time 5203 Seconds  
(without alignments)  
8087.165 Million cell updates/sec

Title: US-10-673-885-1  
Perfect score: 7108  
Sequence: 1 taattgtacttgcagaa.....gcttctcactgtgccttt 7108

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues  
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | % Match | Query Length | DB ID | Description        |
|------------|--------|---------|--------------|-------|--------------------|
| 1          | 7108   | 100.0   | 7108         | 8     | ABZ23703 Human pho |
| 2          | 6800.8 | 95.7    | 6897         | 9     | AAD56411 Human rec |
| 3          | 6799.8 | 95.7    | 7225         | 8     | ABA00717 Human KPP |
| 4          | 6799   | 95.7    | 6900         | 12    | ADH41614 Novel hum |
| 5          | 6683.6 | 94.0    | 6903         | 6     | ABT06283 Human NOV |
| 6          | 6683.6 | 94.0    | 6903         | 12    | ADH41624 Novel hum |
| 7          | 6453   | 90.8    | 6994         | 6     | ABT06281 Human NOV |
| 8          | 6453   | 90.8    | 6994         | 12    | ADH41618 Novel hum |
| 9          | 6340   | 89.2    | 6881         | 12    | ADH41612 Novel hum |
| 10         | 5619.4 | 79.1    | 5877         | 9     | AAD56413 Human rec |
| 11         | 2716.2 | 38.2    | 2739         | 12    | ADH41620 Novel hum |
| 12         | 2551   | 35.9    | 2585         | 6     | ABT06282 Human NOV |
| 13         | 2551   | 35.9    | 2585         | 12    | ADH41622 Novel hum |
| 14         | 2205   | 31.0    | 2220         | 12    | ADH41616 Novel hum |
| 15         | 1071.6 | 15.1    | 2309         | 2     | AAQ73786 Partial p |
| 16         | 905.2  | 12.7    | 990          | 9     | AAD56412 Human rec |
| 17         | 532    | 7.5     | 532          | 2     | AAQ73785 Partial p |
| 18         | 493.6  | 6.9     | 1380         | 12    | ADP29100 Human sec |
| 19         | 319.4  | 4.5     | 321          | 2     | AAQ73782 Partial p |
| 20         | 315    | 4.4     | 254366       | 8     | ABZ23704 Human pho |

|   |    |       |     |      |    |          |                    |
|---|----|-------|-----|------|----|----------|--------------------|
| C | 21 | 230.8 | 3.2 | 412  | 9  | AAD56414 | Aad56414 Human rec |
|   | 22 | 195.2 | 2.7 | 3623 | 2  | AAV82821 | Aav82821 Rabbit os |
|   | 23 | 195.2 | 2.7 | 5679 | 2  | AAV04867 | Aav04867 Nucleotid |
|   | 24 | 193.8 | 2.7 | 6329 | 9  | ACC79776 | Acc79776 Human PTP |
|   | 25 | 192.2 | 2.7 | 936  | 12 | ADO04583 | Ado04583 Human HPT |
|   | 26 | 192.2 | 2.7 | 1011 | 12 | ADO04581 | Ado04581 Human HPT |
|   | 27 | 192.2 | 2.7 | 3984 | 12 | ADH48408 | Adh48408 Human KPP |
|   | 28 | 192.2 | 2.7 | 4752 | 2  | AAV04868 | Aav04868 Nucleotid |
|   | 29 | 192.2 | 2.7 | 4815 | 2  | AAV04886 | Aav04886 Nucleotid |
|   | 30 | 192.2 | 2.7 | 5069 | 10 | ADD89070 | Add89070 Encoding  |
|   | 31 | 192.2 | 2.7 | 5113 | 2  | AAT48479 | Aat48479 Human tyr |
|   | 32 | 192.2 | 2.7 | 5153 | 13 | ADP54714 | Adp54714 Human pro |
|   | 33 | 192.2 | 2.7 | 5153 | 13 | ADP24083 | Adp24083 PRO polyP |
|   | 34 | 192.2 | 2.7 | 5415 | 13 | ADK52802 | Adk52802 Drug ther |
|   | 35 | 192.2 | 2.7 | 5415 | 13 | ADS17077 | Adsl7077 Human 207 |
|   | 36 | 192.2 | 2.7 | 6075 | 3  | AAA88866 | Aaa88866 Human pro |
|   | 37 | 192.2 | 2.7 | 6075 | 11 | ADI32040 | Adi32040 Human cDN |
|   | 38 | 192.2 | 2.7 | 6075 | 12 | ADO04579 | Ado04579 Human HPT |
|   | 39 | 192.2 | 2.7 | 6342 | 4  | AAK51954 | Aak51954 Human pol |
|   | 40 | 192.2 | 2.7 | 6632 | 4  | AAI26742 | Aai26742 Human bre |
|   | 41 | 192.2 | 2.7 | 7444 | 6  | AAD32364 | Aad32364 Human lun |
|   | 42 | 192.2 | 2.7 | 7444 | 10 | ADE54130 | Ades4130 Human pro |
|   | 43 | 192.2 | 2.7 | 7879 | 11 | ACN88701 | Acn88701 Breast ca |
|   | 44 | 187.4 | 2.6 | 4799 | 2  | AAV04887 | Aav04887 Nucleotid |
|   | 45 | 184.4 | 2.6 | 6346 | 4  | AAK52938 | Aak52938 Human pol |

## ALIGNMENTS

RESULT 1  
ABZ23703  
ID ABZ23703 standard; cDNA; 7108 BP.

XX AC ABZ23703;

DT 21-MAR-2003 (first entry)

DE Human phosphatase encoding cDNA #SEQ ID 1.

KW Human; phosphatase; enzyme; chromosome 12; colon adenocarcinoma;  
KW placenta; gene therapy; gene; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT 5'UTR 1..105

FT CDS /\*tag= a

FT /\*tag= b 106..6981

FT /\*product= "phosphatase protein"

FT 3'UTR 6982..7108

FT /\*tag= C

XX WO200279452-A2.

XX 10-OCT-2002.

XX 02-APR-2002; 2002WO-US009992.

XX 02-APR-2001; 2001US-00822871.

XX (PEKE ) PE CORP NY.

XX Webster M, Wei M, Di Francesco V, Beasley EM;

XX WPI; 2003-046809/04.

XX P-PSDB; ABP60057.

PT New human phosphatase peptides and nucleic acids encoding the peptides,  
PT useful as models in developing human therapeutic targets, in identifying  
PT therapeutic proteins or modulators of protein activity, and in gene  
PT therapy.

XX PS Claim 4b; Fig 1; 220pp; English.

XX CC The invention relates to an isolated human phosphatase peptide. The human phosphatase peptides and nucleic acids are useful as models for the development of human therapeutic targets and in the identification of therapeutic proteins. They also serve as targets for the development of human therapeutic agents that modulate phosphatase activity in cells and tissues that express the phosphatase. Experimental data indicates that the cDNA is expressed in colon adenocarcinoma and placenta. The proteins can be used to raise or to elicit another immune response, as a reagent in assays designed to determine the levels of the protein in biological fluids, as markers for tissue in which the corresponding protein is preferentially expressed, in the identification of modulators of the peptides, and in pharmacogenomic analysis. The polypeptides and polynucleotides may be used in gene therapy, and as antisense constructs to control phosphatase gene expression in cells, tissue and organisms. The gene encoding the phosphatase of the invention is located on human chromosome 12. The current sequence represents the human phosphatase encoding cDNA

XX SQ Sequence 7108 BP; 2387 A; 1408 C; 1334 G; 1979 T; 0 U; 0 Other;

Query Match 100.0%; Score 7108; DB 8; Length 7108;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 7108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAATTGTACTTGCAGAGAGATCTGCTTTAAATCAATTAATGACGGAACAATTTCTCT 60  
DB 1 TAATTGTACTTGCAGAGAGATCTGCTTTAAATCAATTAATGACGGAACAATTTCTCT 60

QY 61 CTAGAGCCATCAATGATGATCTACTGCTGAAATGTAATTAAGATGATTTCTTATC 120  
DB 61 CTAGAGCCATCAATGATGATCTACTGCTGAAATGTAATTAAGATGATTTCTTATC 120

QY 121 ATTTTCTTTTACTTTTATTTGGACTTTCAGACACACAGGTGATGTTTCCAATGTCGTT 180  
DB 121 ATTTTCTTTTACTTTTATTTGGACTTTCAGACACACAGGTGATGTTTCCAATGTCGTT 180

QY 181 CTGGTACTAGGTACGATATAACCATCTCTTCAATTTTACAAATACACCTCAGCTGTT 240  
DB 181 CTGGTACTAGGTACGATATAACCATCTCTTCAATTTTACAAATACACCTCAGCTGTT 240

QY 241 ACTAGAAATAGTGACCAAAATGTAACAAACACAGGCGCTCCAGTCTTCTAGCCGGGAA 300  
DB 241 ACTAGAAATAGTGACCAAAATGTAACAAACACAGGCGCTCCAGTCTTCTAGCCGGGAA 300

QY 301 AGAGTCGGATCTGCTGGGATTTCTGCTTGGAAATACACCACTAATCCAAATGGAAGG 360  
DB 301 AGAGTCGGATCTGCTGGGATTTCTGCTTGGAAATACACCACTAATCCAAATGGAAGG 360

QY 361 ATTATATCTTACATTTGTAAGAAATTAAGAAATTTGTCGGTGGATGCAAAACAGTATATACA 420  
DB 361 ATTATATCTTACATTTGTAAGAAATTAAGAAATTTGTCGGTGGATGCAAAACAGTATATACA 420

QY 421 CAAAGTCAGATCAAGCCAGACAGTCTGGAAGTTTCTTCTTACTAATCTTAACTCTGGAACA 480  
DB 421 CAAAGTCAGATCAAGCCAGACAGTCTGGAAGTTTCTTCTTACTAATCTTAACTCTGGAACA 480

QY 481 ACATATGAAATTAAGGTTTGTGCTGAAAAACAGTCTGGCATTTGGAGTGTGATGATCCA 540  
DB 481 ACATATGAAATTAAGGTTTGTGCTGAAAAACAGTCTGGCATTTGGAGTGTGATGATCCA 540

QY 541 TTTTCTCTTCAAACTGAGAAAGTCTCCAGGAAAGTGGTGAATCTCAGTTGAGGCC 600  
DB 541 TTTTCTCTTCAAACTGAGAAAGTCTCCAGGAAAGTGGTGAATCTCAGTTGAGGCC 600

QY 601 TACAACGCTTACAGCTTAAGCTGATTTGATTTTACCTCGGCAACCAATGCAAAAT 660  
DB 601 TACAACGCTTACAGCTTAAGCTGATTTGATTTTACCTCGGCAACCAATGCAAAAT 660

QY 661 ACCAGCTTCAAGATTAAGTGTCAAAACATGCCAGAAAGTGGATAGTGAAGATGTCTCA 720  
DB 661 ACCAGCTTCAAGATTAAGTGTCAAAACATGCCAGAAAGTGGATAGTGAAGATGTCTCA 720

DB 661 ACCAGCTTCAAGATTAAGTGTCAAAACATGCCAGAAAGTGGATAGTGAAGATGTCTCA 720

QY 721 ATCAGAGTAGAGGACATTTTGAATGGGAAATTTGCCAGAAATGCAATGAGNATAGTGAATCT 780  
DB 721 ATCAGAGTAGAGGACATTTTGAATGGGAAATTTGCCAGAAATGCAATGAGNATAGTGAATCT 780

QY 781 TTTTATATGAGTACAGCAGCGCTTCTCAACCCCTTGGTAGAGTTTACACCTCCATCGCGT 840  
DB 781 TTTTATATGAGTACAGCAGCGCTTCTCAACCCCTTGGTAGAGTTTACACCTCCATCGCGT 840

QY 841 ACCACACATTCATCAAGCAGCTTGAACAGAAATGAGATCAGCTCTGTGTGGAAAGAGCT 900  
DB 841 ACCACACATTCATCAAGCAGCTTGAACAGAAATGAGATCAGCTCTGTGTGGAAAGAGCT 900

QY 901 ATCAGTTTGTAGTACACACCTTGAACCTTATACATATATCTTTTGAAGTTTCAGCT 960  
DB 901 ATCAGTTTGTAGTACACACCTTGAACCTTATACATATATCTTTTGAAGTTTCAGCT 960

QY 961 GCTCAACCTGAAGCAGGTTTATATGATAGTACGATTTGTACAGAAACACCAAGATCAGTGCCT 1020  
DB 961 GCTCAACCTGAAGCAGGTTTATATGATAGTACGATTTGTACAGAAACACCAAGATCAGTGCCT 1020

QY 1021 GAAGGACCAACCAAACTGCGTAAACAGGCAACATCACAGGAAAGTCTCTTTCAATTTTA 1080  
DB 1021 GAAGGACCAACCAAACTGCGTAAACAGGCAACATCACAGGAAAGTCTCTTTCAATTTTA 1080

QY 1081 TGGGACCCCAACCACTATAGTAAACAGGAAATTTAGTTATAGAGTTTGAATTTATGAGCA 1140  
DB 1081 TGGGACCCCAACCACTATAGTAAACAGGAAATTTAGTTATAGAGTTTGAATTTATGAGCA 1140

QY 1141 TCAGGTGCGATTTTGGATAACAGCAACAAAGACCTCAAGTTTGGCAATTCATAACCTAAACA 1200  
DB 1141 TCAGGTGCGATTTTGGATAACAGCAACAAAGACCTCAAGTTTGGCAATTCATAACCTAAACA 1200

QY 1201 CCATTTACAATGATGATGTCTATATTTGGGCTGAAACCAAGTGCAGGAGCTGGGCGCAAG 1260  
DB 1201 CCATTTACAATGATGATGTCTATATTTGGGCTGAAACCAAGTGCAGGAGCTGGGCGCAAG 1260

QY 1261 TCAATATTTTCAATTTCTCTCCACAGATGTTTCCAGGGGAGCTTTGATTTTACAATTT 1320  
DB 1261 TCAATATTTTCAATTTCTCTCCACAGATGTTTCCAGGGGAGCTTTGATTTTACAATTT 1320

QY 1321 GCAGAGGTAGAAATCCACGCAAGTAAAGAAATTTCTTGGAAAGAAACACGACCAACCAATGGA 1380  
DB 1321 GCAGAGGTAGAAATCCACGCAAGTAAAGAAATTTCTTGGAAAGAAACACGACCAACCAATGGA 1380

QY 1381 ATTATTAACCAATAACGAGTGAAGTGTAGTTTCCAGAGACAGGAAATTTTGGAAAT 1440  
DB 1381 ATTATTAACCAATAACGAGTGAAGTGTAGTTTCCAGAGACAGGAAATTTTGGAAAT 1440

QY 1441 ACTTTGCTCAGCTGGAATTAAGTATATAAATGACCCCATGCTCCAGAAATTTGTGAAC 1500  
DB 1441 ACTTTGCTCAGCTGGAATTAAGTATATAAATGACCCCATGCTCCAGAAATTTGTGAAC 1500

QY 1501 ATAGTAGAGCCCAATGGTAGGATTTATAGGGTTTCCAGAGAGATGTCGTCTGACCTTCAAC 1560  
DB 1501 ATAGTAGAGCCCAATGGTAGGATTTATAGGGTTTCCAGAGAGATGTCGTCTGACCTTCAAC 1560

QY 1561 TCACTTGTCTACATTTATATAAAGCCATCCAGATAAATACTTCTGCAAGGAATAGA 1620  
DB 1561 TCACTTGTCTACATTTATATAAAGCCATCCAGATAAATACTTCTGCAAGGAATAGA 1620

QY 1621 GCTGAAGACCCAGACTTCCAGCTTGTACTACAAGGAATCAGTATATCTACATGCA 1680  
DB 1621 GCTGAAGACCCAGACTTCCAGCTTGTACTACAAGGAATCAGTATATCTACATGCA 1680

QY 1681 GCTGAAGACCTGCTTATGTTATCAGGAGACTTGTACCTTTTCACTGAGCAGATGAT 1740  
DB 1681 GCTGAAGACCTGCTTATGTTATCAGGAGACTTGTACCTTTTCACTGAGCAGATGAT 1740

QY 1741 GTATCTGCTTTCATCATATGAGGAGGACCAACCAAGTCTCAGTGTAGGACACGT 1800  
DB 1741 GTATCTGCTTTCATCATATGAGGAGGACCAACCAAGTCTCAGTGTAGGACACGT 1800



QY 1801 CAGCAAGTCCAGGCTCATTAAATTTATATAAATTTATATAAATTTATAGTCTTCTCATCTATT 1860  
DB 1801 CAGCAAGTCCAGGCTCATTAAATTTATATAAATTTATATAAATTTATAGTCTTCTCATCTATT 1860  
QY 1861 TTGTTATATTGGGATCTCCAGATATCCCAATGGAAATTAACCTCACTATACGATTTAT 1920  
DB 1861 TTGTTATATTGGGATCTCCAGATATCCCAATGGAAATTAACCTCACTATACGATTTAT 1920  
QY 1921 GCAATGGAATTTGGATACAAACAGAGCATTTCCAGATAACTACCATAGATAACAGCTTCTC 1980  
DB 1921 GCAATGGAATTTGGATACAAACAGAGCATTTCCAGATAACTACCATAGATAACAGCTTCTC 1980  
QY 1981 ATACACGGTTAAAGAAATACACAAATACAAATAGAGTGGCAGGCTCAACCCAGAT 2040  
DB 1981 ATACACGGTTAAAGAAATACACAAATACAAATAGAGTGGCAGGCTCAACCCAGAT 2040  
QY 2041 GGAGAAAGTTCTTTGCTGGAAGAAATGACATCTTTGAGAACTTCAGAGATGAACCG 2100  
DB 2041 GGAGAAAGTTCTTTGCTGGAAGAAATGACATCTTTGAGAACTTCAGAGATGAACCG 2100  
QY 2101 GAATCATCACTCAAGATGTGGAATTAATGATGTTTACCCAGATGAATAAGTTGAAG 2160  
DB 2101 GAATCATCACTCAAGATGTGGAATTAATGATGTTTACCCAGATGAATAAGTTGAAG 2160  
QY 2161 TGGTCAACCCGAAAGCCCAATGGATCAATTATTGCTTATGAAGTCTTATATAAAAT 2220  
DB 2161 TGGTCAACCCGAAAGCCCAATGGATCAATTATTGCTTATGAAGTCTTATATAAAAT 2220  
QY 2221 ATAGATCTTTATATAGAGACACATCAACACAGACATATATTAAGGAATTAAGA 2280  
DB 2221 ATAGATCTTTATATAGAGACACATCAACACAGACATATATTAAGGAATTTAAGA 2280  
QY 2281 CCTCACACCTCTATAACATTTCTGTAAGTCTTACACAGATTTGCTCATGCAATCAG 2340  
DB 2281 CCTCACACCTCTATAACATTTCTGTAAGTCTTACACAGATTTGCTCATGCAATCAG 2340  
QY 2341 GTATCTTTTACTCTCTGTAAGGACTTCGGAGACTGTGCTGATAGTGCACAGAAAT 2400  
DB 2341 GTATCTTTTACTCTCTGTAAGGACTTCGGAGACTGTGCTGATAGTGCACAGAAAT 2400  
QY 2401 ATCACTTACAAATATTTCTCTGGAGACTTGAGCTATCATCTCTCCCAAGTAGT 2460  
DB 2401 ATCACTTACAAATATTTCTCTGGAGACTTGAGCTATCATCTCTCCCAAGTAGT 2460  
QY 2461 CCCAATGGAATCATAAAAATATACAAATTTATCTCAAGAGAGTAAATGGAATGAGAA 2520  
DB 2461 CCCAATGGAATCATAAAAATATACAAATTTATCTCAAGAGAGTAAATGGAATGAGAA 2520  
QY 2521 AGAATATAAATCAACCTCTTTAAACCCAAACATTAAGTACTGAAGAAATATACCCAA 2580  
DB 2521 AGAATATAAATCAACCTCTTTAAACCCAAACATTAAGTACTGAAGAAATATACCCAA 2580  
QY 2581 TATATCATTTGAGTGTCTAGTACACTGAAAGGTGAAGGTTGCGAGTGTCCCATTA 2640  
DB 2581 TATATCATTTGAGTGTCTAGTACACTGAAAGGTGAAGGTTGCGAGTGTCCCATTA 2640  
QY 2641 AGTATCTGACGAGGAGAGTGTCTGATTTCTCCCTCAAGCTTCTCTGTAAGACAG 2700  
DB 2641 AGTATCTGACGAGGAGAGTGTCTGATTTCTCCCTCAAGCTTCTCTGTAAGACAG 2700  
QY 2701 TTGCTGGTGTACGGTGAAGTGTCTATGGCAACACCCCTGGAGCCAAATGGAATATC 2760  
DB 2701 TTGCTGGTGTACGGTGAAGTGTCTATGGCAACACCCCTGGAGCCAAATGGAATATC 2760  
QY 2761 CTTTATTAACAGTTTATGTCTGGAATAGATCATCATTAATAAATTAATGTCACCTGAA 2820  
DB 2761 CTTTATTAACAGTTTATGTCTGGAATAGATCATCATTAATAAATTAATGTCACCTGAA 2820  
QY 2821 ACATCATTTGAGTTATCAGATTTGGATTAATGTTGAATACAGTGTCTATGTAACAGCT 2880  
DB 2821 ACATCATTTGAGTTATCAGATTTGGATTAATGTTGAATACAGTGTCTATGTAACAGCT 2880

QY 2881 AGCACAGATTTGGTGAATGGAAAAACAGGAAGCAATATCATTTAGCTTTCAAAACACAGAG 2940  
DB 2881 AGCACAGATTTGGTGAATGGAAAAACAGGAAGCAATATCATTTAGCTTTCAAAACACAGAG 2940  
QY 2941 GGAGACCAAGCGATCCTCCCAAGATGTTTATATGCAAACTCAGTTCTTCTCATCAATA 3000  
DB 2941 GGAGACCAAGCGATCCTCCCAAGATGTTTATATGCAAACTCAGTTCTTCTCATCAATA 3000  
QY 3001 ATTCTTTCTGGACACCTCTTCAAAACCTTAATGGGATTTATACAATATTACTCTGTTAT 3060  
DB 3001 ATTCTTTCTGGACACCTCTTCAAAACCTTAATGGGATTTATACAATATTACTCTGTTAT 3060  
QY 3061 TACAGAAATACCTTCAAGTACTTTTATGCAAACTTATGCACTCCATGAACTAAACCAATGAC 3120  
DB 3061 TACAGAAATACCTTCAAGTACTTTTATGCAAACTTATGCACTCCATGAACTAAACCAATGAC 3120  
QY 3121 TTTGACAAATGACTGTATCCCAATATATAGATAAATGACATATTTGAGTCTACTATACA 3180  
DB 3121 TTTGACAAATGACTGTATCCCAATATATAGATAAATGACATATTTGAGTCTACTATACA 3180  
QY 3181 TTTTGGTTAACCAAGTACTTTCAGTTGGAATGGGAATAAAAAGCAGTGACATCATTTGAA 3240  
DB 3181 TTTTGGTTAACCAAGTACTTTCAGTTGGAATGGGAATAAAAAGCAGTGACATCATTTGAA 3240  
QY 3241 GTATACACAGATCAAGACATACCTGAAAGGTTTGTGGAAACCTGACTTACGAATCCATT 3300  
DB 3241 GTATACACAGATCAAGACATACCTGAAAGGTTTGTGGAAACCTGACTTACGAATCCATT 3300  
QY 3301 TCGTCAACTGCAATTAATGTAAGTGGTCCCAACCGGCTCAACCAACGGTCTAGTCTTC 3360  
DB 3301 TCGTCAACTGCAATTAATGTAAGTGGTCCCAACCGGCTCAACCAACGGTCTAGTCTTC 3360  
QY 3361 TACTATGTTTCACTGATCTTACAGACAGCTCTCGCCATGTGAGACACCTCTTGTATACA 3420  
DB 3361 TACTATGTTTCACTGATCTTACAGACAGCTCTCGCCATGTGAGACACCTCTTGTATACA 3420  
QY 3421 TATGAGAGAGCATATATTTGATATCTGAAAAATACACTGATTTATATATTAATAAAT 3480  
DB 3421 TATGAGAGAGCATATATTTGATATCTGAAAAATACACTGATTTATATATTAATAAAT 3480  
QY 3481 ACTCCATCAACAGAAAGGATTTCTGATACCTATACCTGATGCTCCAGCTATACATCAAGCT 3540  
DB 3481 ACTCCATCAACAGAAAGGATTTCTGATACCTATACCTGATGCTCCAGCTATACATCAAGCT 3540  
QY 3541 GAAGAAGATGCTCCAGAAACTTCAACCAATTAACACACTTTTAAAAAACCCTTCTCTACC 3600  
DB 3541 GAAGAAGATGCTCCAGAAACTTCAACCAATTAACACACTTTTAAAAAACCCTTCTCTACC 3600  
QY 3601 TCAGTTCTCTTATCATGGGATCCCCAGTAAAGCCAAATGGTGCATATAATAAGTTATGAT 3660  
DB 3601 TCAGTTCTCTTATCATGGGATCCCCAGTAAAGCCAAATGGTGCATATAATAAGTTATGAT 3660  
QY 3661 TTAACCTTACAGGACCAATGAAATTTATCTTTTCATTTACTCTGATTAATTACATAATA 3720  
DB 3661 TTAACCTTACAGGACCAATGAAATTTATCTTTTCATTTACTCTGATTAATTACATAATA 3720  
QY 3721 TTGGAAGAGCTTTCACCTTATATATAGCTTTTGTGCGCAAGAACTAGAAAA 3780  
DB 3721 TTGGAAGAGCTTTCACCTTATATATAGCTTTTGTGCGCAAGAACTAGAAAA 3780  
QY 3781 GGACTTGGTCTTCCAGTATTTCTTTTACACAGATGAGTCACTGCGTTCAGCT 3840  
DB 3781 GGACTTGGTCTTCCAGTATTTCTTTTACACAGATGAGTCACTGCGTTCAGCT 3840  
QY 3841 CCAAAAAATTTGACTTTTAACTGATCTTGAATGGTGAATGGAGCCCA 3900  
DB 3841 CCAAAAAATTTGACTTTTAACTGATCTTGAATGGTGAATGGAGCCCA 3900  
QY 3901 AGTCTCTTCCAGGTTGTTTAAAGTATATAGTTTAAATTCATGAACATGAACCT 3960  
DB 3901 AGTCTCTTCCAGGTTGTTTAAAGTATATAGTTTAAATTCATGAACATGAACCT 3960  
QY 3961 GACACTATATATATAAGAAATATATACAGGATTTAAAACTGAAGCCAAACCTTGTGACCTG 4020

|      |    |  |  |      |
|------|----|--|--|------|
| 3961 | Db |  | GACACTATATATATATAAGAAATATATCAGGATTTAAAACTGAAGCCAAACTGTTGTGGACTG  | 4020 |
| 4021 | Qy |  | GAACCACTGACGACCTACTCTATCCGGTGATCTCGCGTTACCAAAAGTTGGAAATGCAAT     | 4080 |
| 4021 | Db |  | GAACCACTGACGACCTACTCTATCCGGTGATCTCGCGTTACCAAAAGTTGGAAATGCAAT     | 4080 |
| 4081 | Qy |  | CAATTTAGTAAATGTAGTAAAAATTCACAACCAAGAATCAGTTCAGAGTGTGTCGACAGAT    | 4140 |
| 4081 | Db |  | CAATTTAGTAAATGTAGTAAAAATTCACAACCAAGAATCAGTTCAGAGTGTGTCGACAGAT    | 4140 |
| 4141 | Qy |  | ATGCAGTGCATGGCAACTAGCTGGCAGTCAAGTTTTAGTGAAATGGGAATCCACCAAAAAG    | 4200 |
| 4141 | Db |  | ATGCAGTGCATGGCAACTAGCTGGCAGTCAAGTTTTAGTGAAATGGGAATCCACCAAAAAG    | 4200 |
| 4201 | Qy |  | GCAAAATGGAATATAAGCGCAGTATATGGTAAACAGTTGAAAGGAATTCACAAAAGTTTCT    | 4260 |
| 4201 | Db |  | GCAAAATGGAATATAAGCGCAGTATATGGTAAACAGTTGAAAGGAATTCACAAAAGTTTCT    | 4260 |
| 4261 | Qy |  | CCCCAAGATCACATGTACACTTTTCATAAAGCTTCTGCCAATACCTCATATGTCTTTAAA     | 4320 |
| 4261 | Db |  | CCCCAAGATCACATGTACACTTTTCATAAAGCTTCTGCCAATACCTCATATGTCTTTAAA     | 4320 |
| 4321 | Qy |  | GTAAGAGCTTCAACCTCAGCTGGTGGAAGGTGATGAAGACACATGCCATGTCAGCACATA     | 4380 |
| 4321 | Db |  | GTAAGAGCTTCAACCTCAGCTGGTGGAAGGTGATGAAGACACATGCCATGTCAGCACATA     | 4380 |
| 4381 | Qy |  | CCTGAAACAGTTCACAGTGTCCCACAAAATATTTGCTTTTTCTGATGTTCACTCAACTAGT    | 4440 |
| 4381 | Db |  | CCTGAAACAGTTCACAGTGTCCCACAAAATATTTGCTTTTTCTGATGTTCACTCAACTAGT    | 4440 |
| 4441 | Qy |  | GCAACATTGACATGGAATAGACCTGACACTATCCTTGGCTACTTTTCAAAATTCAAAAAT     | 4500 |
| 4441 | Db |  | GCAACATTGACATGGAATAGACCTGACACTATCCTTGGCTACTTTTCAAAATTCAAAAAT     | 4500 |
| 4501 | Qy |  | ACCACTCAACTTCGTGCTCAAAAATGCAAAAGATGGGAATCCGAAGAATGTGTGAATAT      | 4560 |
| 4501 | Db |  | ACCACTCAACTTCGTGCTCAAAAATGCAAAAGATGGGAATCCGAAGAATGTGTGAATAT      | 4560 |
| 4561 | Qy |  | CAAAAAATTCAACTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGATGATTAAG       | 4620 |
| 4561 | Db |  | CAAAAAATTCAACTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGATGATTAAG       | 4620 |
| 4621 | Qy |  | AAATTTAGATGGTATAGATTCCAAGTGCCTGCCAGCACCAATGCTGGCTATGCAATGCT      | 4680 |
| 4621 | Db |  | AAATTTAGATGGTATAGATTCCAAGTGCCTGCCAGCACCAATGCTGGCTATGCAATGCT      | 4680 |
| 4681 | Qy |  | TCAACTCGGATTTCTACAAAACTCTGCCTGGCCCTCCAGATGGTCTCTCGTGAATGTT       | 4740 |
| 4681 | Db |  | TCAACTCGGATTTCTACAAAACTCTGCCTGGCCCTCCAGATGGTCTCTCGTGAATGTT       | 4740 |
| 4741 | Qy |  | CATGTAGTAGCAACATCACTTTTATAGCATCAGCATAGAGTGGAGTGAACCTGCTGTCATT    | 4800 |
| 4741 | Db |  | CATGTAGTAGCAACATCACTTTTATAGCATCAGCATAGAGTGGAGTGAACCTGCTGTCATT    | 4800 |
| 4801 | Qy |  | ACTGGACCACTGTTATCTGATTTGATGTCAAAATCGGTAGATATGATGATTAATATA        | 4860 |
| 4801 | Db |  | ACTGGACCACTGTTATCTGATTTGATGTCAAAATCGGTAGATATGATGATTAATATA        | 4860 |
| 4861 | Qy |  | TCCTTTCAATCAAGTCAAAATGAAGAAAAATAAACCACTAGAAAATTAAGAAATTTAGAAAT   | 4920 |
| 4861 | Db |  | TCCTTTCAATCAAGTCAAAATGAAGAAAAATAAACCACTAGAAAATTAAGAAATTTAGAAAT   | 4920 |
| 4921 | Qy |  | ACAAAGTATTTCTGTAGTGAATCACTGCAATTTATCTGGGAACAATTAGTGTGCAATATGAGAA | 4980 |
| 4921 | Db |  | ACAAAGTATTTCTGTAGTGAATCACTGCAATTTATCTGGGAACAATTAGTGTGCAATATGAGAA | 4980 |
| 4981 | Qy |  | GGGAAGTCAAGTGTGAAATGATTTGTTACTTTTACTGAAATCGAGTGAAGTCAAAATTTCAAT  | 5040 |
| 4981 | Db |  | GGGAAGTCAAGTGTGAAATGATTTGTTACTTTTACTGAAATCGAGTGAAGTCAAAATTTCAAT  | 5040 |
| 5041 | Qy |  | AACAACATGCACTTTACAGAGATACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT     | 5100 |

|      |    |   |      |
|------|----|---|------|
| 5041 | Db | AAACAATGACATTTTCAGAGATACCAAGATGAAGTTTCAAAAATTTCAATTTAAAGTTCCCTT     | 5101 |
| 5101 | Qy | CCTCCTTCTCAAACTTAATGGAATATATCCAAAGTATATCAAGCTCTGGTTTACCGAGAAGAT     | 5160 |
|      | Db |   |      |
|      | Db | CCTCCTTCTCAACCTTAATGGAATATATCCAAAGTATATCAAGCTCTGGTTTACCGAGAAGAT     | 5160 |
| 5161 | Qy | GATCCTACTGCTGTCAGATTTCAACAACCTCAGTATTATATACAGAAAAACCAACATTCGTC      | 5220 |
|      | Db | GATCCTACTGCTGTCAGATTTCAACAACCTCAGTATTATATACAGAAAAACCAACATTCGTC      | 5220 |
|      | Qy | ATTTGCAATGCTAGNAGCACTTAAAGGTGGACATACATACAAATCATCAGTGTTTACGCACTG     | 5280 |
| 5281 | Db | ATTTGCAATGCTAGNAGCACTTAAAGGTGGACATACATACAAATCATCAGTGTTTACGCACTG     | 5280 |
|      | Qy | AATAGTCTGCTGCAGGTCCAAAGTTTCCGATCAGAAATAACCATGGATATCAAAAGCTCCA       | 5340 |
|      | Db |   |      |
| 5341 | Qy | ATATAGTCTGCTGCAGGTCCAAAGTTTCCGATCAGAAATAACCATGGATATCAAAAGCTCCA      | 5340 |
|      | Db |   |      |
|      | Qy | GCACGACCAAAAAACCAAAACCAACCCCTTATTTATGATGCCACAGGAAAACTGCTTGTGACT     | 5400 |
| 5401 | Db | GCACGACCAAAAAACCAAAACCAACCCCTTATTTATGATGCCACAGGAAAACTGCTTGTGACT     | 5400 |
|      | Qy | TCAACAACAATTTACAAATCAGAAATGCCAAATATGTTACTACAGTGATGATCATGAGCAACATA   | 5460 |
|      | Db | TCAACAACAATTTACAAATCAGAAATGCCAAATATGTTACTACAGTGATGATCATGAGCAACATA   | 5460 |
| 5461 | Qy | AAAAATGTACAAGTCTTGGCAGACAAACAGGAGCTCAGCATGATGAAAAATGTAAACAAG        | 5520 |
|      | Db | AAAAATGTACAAGTCTTGGCAGACAAACAGGAGCTCAGCATGATGAAAAATGTAAACAAG        | 5520 |
|      | Qy | TGGTATGATGCATATTTTAAATTAAGCAAGGCCATATTTTACAAATGAAGGCTTTTCTTAAC      | 5580 |
| 5581 | Db | TGGTATGATGCATATTTTAAATTAAGCAAGGCCATATTTTACAAATGAAGGCTTTTCTTAAC      | 5580 |
|      | Qy | CCTCCATGTACAGAAGAAAGACAAAGTTTATGAGGCAATGAAGAAATCTCATCATCATAGT       | 5640 |
|      | Db | CCTCCATGTACAGAAGAAAGACAAAGTTTATGAGGCAATGAAGAAATCTCATCATCATAGT       | 5640 |
| 5641 | Qy | GCTCATATGATCATGATTTTAAATTAAGCAAGGCCATATTTTACAAATGAAGGCTTTTCTTAAC    | 5700 |
|      | Db | GCTCATATGATCATGATTTTAAATTAAGCAAGGCCATATTTTACAAATGAAGGCTTTTCTTAAC    | 5700 |
|      | Qy | CCAAAAAGCAATCTTATTTAAATTTAGAGCTACAAATATTTAGGCAAAATTTTACGAC          | 5760 |
| 5761 | Db | CCAAAAAGCAATCTTATTTAAATTTAGAGCTACAAATATTTAGGCAAAATTTTACGAC          | 5760 |
|      | Qy | TCTGATTTATCTGACCCCTGTTTAAAGCTTTAGGGGAAGCACTTTCAGAAAAACCGTAGAG       | 5820 |
|      | Db | TCTGATTTATCTGACCCCTGTTTAAAGCTTTAGGGGAAGCACTTTCAGAAAAACCGTAGAG       | 5820 |
| 5821 | Qy | ATCATTTCTTCCGTCACCTTTGTTGATCCTTTTCAATAATTTCTCCTTGGAAACAGCTATTTTT    | 5880 |
|      | Db | ATCATTTCTTCCGTCACCTTTGTTGATCCTTTTCAATAATTTCTCCTTGGAAACAGCTATTTTT    | 5880 |
|      | Qy | ATCATTTCTTCCGTCACCTTTGTTGATCCTTTTCAATAATTTCTCCTTGGAAACAGCTATTTTT    | 5880 |
| 5881 | Db | ATCATTTCTTCCGTCACCTTTGTTGATCCTTTTCAATAATTTCTCCTTGGAAACAGCTATTTTT    | 5880 |
|      | Qy | GCATTTGCAAGAAATTCGACAGAAAGCAAGAAAGAGGTGGCACAATCTCTCCTCAGGATGCA      | 5940 |
|      | Db | GCATTTGCAAGAAATTCGACAGAAAGCAAGAAAGAGGTGGCACAATCTCTCCTCAGGATGCA      | 5940 |
| 5941 | Qy | GAAATTTATGACATAAATTTGAAGCTGGATCAGCTCATCATCAGTGGCAGACCTGGAACTG       | 6000 |
|      | Db | GAAATTTATGACATAAATTTGAAGCTGGATCAGCTCATCATCAGTGGCAGACCTGGAACTG       | 6000 |
|      | Qy | AAGGACGAGAGATTTAAACCGGCCAATTAAGCAAGAAATTCCTTCTTCGCAACATGTTGAAGAG    | 6060 |
| 6061 | Db | AAGGACGAGAGATTTAAACCGGCCAATTAAGCAAGAAATTCCTTCTTCGCAACATGTTGAAGAG    | 6060 |
|      | Qy | CTTTTGCACAAAACCAACCTTAAAGCTTTCAAGAAAGAAATTTTTCGGAATTTTACCAAAATTTCTT | 6120 |
|      | Db | CTTTTGCACAAAACCAACCTTAAAGCTTTCAAGAAAGAAATTTTTCGGAATTTTACCAAAATTTCTT | 6120 |
| 6121 | Qy | CAGGATCTTCTTCAACTGATGCTGATCTGCTTTGGAAATAGAGCAAAAAACCGTTTCCCA        | 6180 |
|      | Db | CAGGATCTTCTTCAACTGATGCTGATCTGCTTTGGAAATAGAGCAAAAAACCGTTTCCCA        | 6180 |



QY 346 AATCCAAATGGAAGGATTATCTTACATTTGTCAAATATAAGGAAGTTTGTCCGTGGATG 405  
DB 241 AATCCAAATGGAAGGATTATCTTACATTTGTCAAATATAAGGAAGTTTGTCCGTGGATG 300  
QY 406 CAAACAGTATATACACAGTCAGATCAAGACCCAGACAGCTCTGGAAGTTCTTCTTACTAAT 465  
DB 301 CAAACAGTATATACACAGTCAGATCAAGACCCAGACAGCTCTGGAAGTTCTTCTTACTAAT 360  
QY 466 CTTTAATCTGGAAACAATATGAATTAAGTTTGTCTGTGAAACAAGCTGTGGCATTTGGA 525  
DB 361 CTTTAATCTGGAAACAATATGAATTAAGTTTGTCTGTGAAACAAGCTGTGGCATTTGGA 420  
QY 526 GTCTTTAGTGATCCATTTCTCTTCCAAATGTCAGAAAGTGCTCCAGGAAAGTGGTGAAT 585  
DB 421 GTCTTTAGTGATCCATTTCTCTTCCAAATGTCAGAAAGTGCTCCAGGAAAGTGGTGAAT 480  
QY 586 CTCACAGTTGAGCCCTACAAAGCTTCAGCAGTTTAAGCTGATTTGGTATTTTACCTCGGCAA 645  
DB 481 CTCACAGTTGAGCCCTACAAAGCTTCAGCAGTTTAAGCTGATTTGGTATTTTACCTCGGCAA 540  
QY 646 CGAAATGGCAAAATTAACAGCTTCAAGATTAGTGTCAACATGCCAGAAAGTGGATAGTA 705  
DB 541 CGAAATGGCAAAATTAACAGCTTCAAGATTAGTGTCAACATGCCAGAAAGTGGATAGTA 600  
QY 706 GTGAAAGATGTCTCAATCAGAGTAGAGGACATTTTGACTGGGAAATTTGCCAGAAATGCAAT 765  
DB 601 GTGAAAGATGTCTCAATCAGAGTAGAGGACATTTTGACTGGGAAATTTGCCAGAAATGCAAT 660  
QY 766 GAGAATAGTGAATCTTTTATGGAGTACAGCCAGCCCTTCTCCAAACCTTTGGTAGATT 825  
DB 661 GAGAATAGTGAATCTTTTATGGAGTACAGCCAGCCCTTCTCCAAACCTTTGGTAGATT 720  
QY 826 ACACCTCCATCGGTGACACATTCATCATAGCAGCTTTGACACAGAAATGAGATCAGCTCT 885  
DB 721 ACACCTCCATCGGTGACACATTCATCATAGCAGCTTTGACACAGAAATGAGATCAGCTCT 780  
QY 886 GTGTGAAAGAGCCTATCAGTTTGTAGTGACACCTTGAGACCTTTATCAACATATCTT 945  
DB 781 GTGTGAAAGAGCCTATCAGTTTGTAGTGACACCTTGAGACCTTTATCAACATATCTT 840  
QY 946 TTTGAAAGTTTCAGCTGCTACAACTGAAGCAGGTTATATATGATAGTACGATTTGTGAGAAC 1005  
DB 841 TTTGAAAGTTTCAGCTGCTACAACTGAAGCAGGTTATATGATAGTACGATTTGTGAGAAC 900  
QY 1006 CAGAAATCAGTGCTGAGAGACACCAAACTGGTGTACAGGCAACATCAGAGAAAG 1065  
DB 901 CCAGAAATCAGTGCTGAGAGACACCAAACTGGTGTACAGGCAACATCAGAGAAAG 960  
QY 1066 TCCCTTTTCAATTTTATGGGACCCACCACTATATAGTAAAGGAAATTTTATGTTATAGATT 1125  
DB 961 TCCCTTTTCAATTTTATGGGACCCACCACTATATAGTAAAGGAAATTTTATGTTATAGATT 1020  
QY 1126 GAAATATATGGACCATCAGTCTGCATTTTGGATTAACAGCAAAAGACCTCAAGTTTGCA 1185  
DB 1021 GAAATATATGGACCATCAGTCTGCATTTTGGATTAACAGCAAAAGACCTCAAGTTTGCA 1080  
QY 1186 TTCACTAACTTAACACATTTTACAAATGTATGTCTATATTTGGCTGTAAACCAAGTGTGCA 1245  
DB 1081 TTCACTAACTTAACACATTTTACAAATGTATGTCTATATTTGGCTGTAAACCAAGTGTGCA 1140  
QY 1246 GGGACTGGGCCCAGTCAAAATTTTCCAGTATTCACCTCCACAGATGTTCCAGGGGAGTG 1305  
DB 1141 GGGACTGGGCCCAGTCAAAATTTTCCAGTATTCACCTCCACAGATGTTCCAGGGGAGTG 1200  
QY 1306 TTTGATTTTCAACTTTCAGAGGTAGAAATCCACCAAGTAAAGTATCTTGGAGAAACCA 1365  
DB 1201 TTTGATTTTCAACTTTCAGAGGTAGAAATCCACCAAGTAAAGTATCTTGGAGAAACCA 1260  
QY 1366 CGAACCAAAATGGAATTTTAAACCAATACCGAGTGAAAGTGTAGTTCCAGAGACAGGA 1425  
DB 1261 CGAACCAAAATGGAATTTTAAACCAATACCGAGTGAAAGTGTAGTTCCAGAGACAGGA 1320

QY 1426 ATAAATTTTGGAAATATCTTTGCTCACTGGAATAATAGATATATAAATGACCCCATGGCT 1485  
DB 1321 ATAAATTTTGGAAATATCTTTGCTCACTGGAATAATAGATATATAAATGACCCCATGGCT 1380  
QY 1486 CCAGAAATTTGTGAAACATAGTAGAGCAATGGTAGGATATATAGGGTTTCAGCAGAGATG 1545  
DB 1381 CCAGAAATTTGTGAAACATAGTAGAGCAATGGTAGGATATATAGGGTTTCAGCAGAGATG 1440  
QY 1546 TCGTCTGTGACCTTCACTCACTTCTGCTACATTTATATATAACAGCCATCCAGATAAAACTTTT 1605  
DB 1441 TCGTCTGTGACCTTCACTCACTTCTGCTACATTTATATATAACAGCCATCCAGATAAAACTTTT 1500  
QY 1606 CTTGCAAGGAATAGAGCTGAAGACAGACTTCAACAGTTGTAACTTACAGGAATCAGTAT 1665  
DB 1501 CTTGCAAGGAATAGAGCTGAAGACAGACTTCAACAGTTGTAACTTACAGGAATCAGTAT 1560  
QY 1666 ATTACTGACATTTGACAGCTGAAACAGCTGCTTAAGTTTACAGAGACTTGTACTTTCACT 1725  
DB 1561 ATTACTGACATTTGACAGCTGAAACAGCTGCTTATGTATTCAGAGACTTGTACTTTCACT 1620  
QY 1726 GAGCACATGATTAGTGTATCTGCTTTCACCATCATGGGGAAGGACCAACACAGTTCTC 1785  
DB 1621 GAGCACATGATTAGTGTATCTGCTTTCACCATCATGGGGAAGGACCAACACAGTTCTC 1680  
QY 1786 AGTGTTAGGACACGTGACCAAGCTCCATTTAAATTTATAAACTATAAAATATTT 1845  
DB 1681 AGTGTTAGGACACGTGACCAAGCTCCATTTAAATTTATAAACTATAAAATATTT 1740  
QY 1846 AGTGTTCATCTATTTTGTATTTTGGGATCTCCAGAAATATCCCAATGGAATAATTAAC 1905  
DB 1741 AGTGTTCATCTATTTTGTATTTTGGGATCTCCAGAAATATCCCAATGGAATAATTAAC 1800  
QY 1906 CACTATACGATTTATGCAATGGAAATTTGGATACAAACAGAGCAATCCAGATAACTACCA 1965  
DB 1801 CACTATACGATTTATGCAATGGAAATTTGGATACAAACAGAGCAATCCAGATAACTACCA 1860  
QY 1966 GATAACAGCTTTCTCATAAACAGGTTTAAAGAAATACAAAAATACAAAATGAGAGTGCA 2025  
DB 1861 GATAACAGCTTTCTCATAAACAGGTTTAAAGAAATACAAAATACAAAATGAGAGTGCA 1920  
QY 2026 GCTTCAACCCACAGATGGGAAAGTTCTTTGTCTGTAAGAAATGACATCTTTGTGAGAACT 2085  
DB 1921 GCTTCAACCCACAGATGGGAAAGTTCTTTGTCTGTAAGAAATGACATCTTTGTGAGAACT 1980  
QY 2086 TCAGAGATGAAACCGGAATCATCACTCAAGATGTGGAATTTGATGTTTACCCGAGAT 2145  
DB 1981 TCAGAGATGAAACCGGAATCATCACTCAAGATGTGGAATTTGATGTTTACCCGAGAT 2040  
QY 2146 GAAATAAGGTTTGAAGTGGTCAACCCCGAAAAAGCCCAATGGGATCATTTATGCTTTATGAA 2205  
DB 2041 GAAATAAGGTTTGAAGTGGTCAACCCCGAAAAAGCCCAATGGGATCATTTATGCTTTATGAA 2100  
QY 2206 GTGCTATATAAAAAATATAGATATCTTTATATATGAAAGACACATCAACACAGACATAATA 2265  
DB 2101 GTGCTATATAAAAAATATAGATATCTTTATATATGAAAGACACATCAACACAGACATAATA 2160  
QY 2266 TTAAGGAACTTAAGACCTCCACACCTCTATAACATTTCTGTAAGGTCTTTACACAGATTT 2325  
DB 2161 TTAAGGAACTTAAGACCTCCACACCTCTATAACATTTCTGTAAGGTCTTTACACAGATTT 2220  
QY 2326 GGTCTATGGCAATCAGGTATCTTTTACTCTCTGTAAGGACCTTCGGAGACTGTGCTGAT 2385  
DB 2221 GGTCTATGGCAATCAGGTATCTTTTACTCTCTGTAAGGACCTTCGGAGACTGTGCTGAT 2280  
QY 2386 AGTGCAACGAAAAATATCACTTACAAAAATATTTCTTCTGGAGAGATTTGAGCTATCATTT 2445  
DB 2281 AGTGCAACGAAAAATATCACTTACAAAAATATTTCTTCTGGAGAGATTTGAGCTATCATTT 2340  
QY 2446 CTTTCCCCCAAGTAGTCCCAATGGAATCATAAAAAATATACAAATTTTATCTCAAGAGAACT 2505  
DB 2341 CTTTCCCCCAAGTAGTCCCAATGGAATCATAAAAAATATACAAATTTTATCTCAAGAGAACT 2400  
QY 2506 AATGGAAATGAGGAAAGAACTATAAATAACAACCTCTTTTAAACCCCAAAACATTAAGTACTG 2565

Db 2401 AATGGAAATGAGGAAAGCACTATAATACAACTCTTTAAACCCAAACCAATTAAGTACTG 2460  
Qy |||||  
Db 2566 AAGAAATATACCAATATATCATTTGAGGTCTCTGCTAGTACACTGAAAGGTGAAGAGTT 2625  
Qy |||||  
Db 2461 AAGAAATATACCAATATATCATTTGAGGTCTCTGCTAGTACACTCAAGGTGAAGAGTT 2520  
Qy |||||  
Db 2626 CGAGTCTCCCATAGATATACGACGAGAGAGATGCTCTGATTTCTCCCCCTCAAGAC 2685  
Qy |||||  
Db 2521 CGAGTCTCCCATAGATATACGACGAGAGAGATGCTCTGATTTCTCCCCCTCAAGAC 2580  
Qy |||||  
Db 2686 TTCTCTGTAAGAACAGTTGCTGCTGTCACGGTGAAGTTGTCATGGCAACACCCCTGGAG 2745  
Qy |||||  
Db 2581 TTCTCTGTAAGAACAGTTGCTGCTGTCACGGTGAAGTTGTCATGGCAACACCCCTGGAG 2640  
Qy |||||  
Db 2746 CCAATGGAATTAATCTTTATTACACAGTTTATGCTGGAATAGATCATATTAATAAAT 2805  
Qy |||||  
Db 2641 CCAATGGAATTAATCTTTATTACACAGTTTATGCTGGAATAGATCATATTAATAAAT 2700  
Qy |||||  
Db 2806 ATTAATGTCCTCAAAATCATCATTTGAGGTATACAGATTTGGATTAATATGTTGAATACAGT 2865  
Qy |||||  
Db 2701 ATTAATGTCCTCAAAATCATCATTTGAGGTATACAGATTTGGATTAATATGTTGAATACAGT 2760  
Qy |||||  
Db 2866 GCTTATGTAACAGCTAGCAGCAATTTGGTATGGGAAACACAGAAACCAATATCATTAGC 2925  
Qy |||||  
Db 2761 GCTTATGTAACAGCTAGCAGCAATTTGGTATGGGAAACACAGAAACCAATATCATTAGC 2820  
Qy |||||  
Db 2926 TTTCAAACACAGAGGGAGCACCAAGCGATCTCTCCAAAGATGTTTATTATGCAAACTC 2985  
Qy |||||  
Db 2821 TTTCAAACACAGAGGGAGCACCAAGCGATCTCTCCAAAGATGTTTATTATGCAAACTC 2880  
Qy |||||  
Db 2986 AGTCTTCTCAATAATCTTTTCTGGACACCTCTCTTCAAACTAATGGATTAACAA 3045  
Qy |||||  
Db 2881 AGTCTTCTCAATAATCTTTTCTGGACACCTCTCTTCAAACTAATGGATTAACAA 2940  
Qy |||||  
Db 3046 TATTACTCTGTTTATTACAGAAATACCTTCAGGTACTTTTATGCGAGATTTTACACTCCAT 3105  
Qy |||||  
Db 2941 TATTACTCTGTTTATTACAGAAATACCTTCAGGTACTTTTATGCGAGATTTTACACTCCAT 3000  
Qy |||||  
Db 3106 GAACCTAACCAATGACCTTGCAATATGACTGTATCCAAATTAAGTAACTGCAATA 3165  
Qy |||||  
Db 3001 GAAGTAACCAATGACCTTGCAATATGACTGTATCCAAATTAAGTAACTGCAATA 3060  
Qy |||||  
Db 3166 TTCAGTCTACTATACATTTTGGTTTAAACAGCAAGTACTTTCAGTTGGAATGGGAATAAAGC 3225  
Qy |||||  
Db 3061 TTCAGTCTACTATACATTTTGGTTTAAACAGCAAGTACTTTCAGTTGGAATGGGAATAAAGC 3120  
Qy |||||  
Db 3226 AGTGACATCATTTGAAGTATACACAGATCAAGATCAATACCTGGAAGGGTTTGGTGAACCTG 3285  
Qy |||||  
Db 3121 AGTGACATCATTTGAAGTATACACAGATCAAGATCAATACCTGGAAGGGTTTGGTGAACCTG 3180  
Qy |||||  
Db 3286 ACTTACGAAATCCATTTTCGTCACCTGCAATTAATGTAAGTGGTCCACCGCTCAACCA 3345  
Qy |||||  
Db 3181 ACTTACGAAATCCATTTTCGTCACCTGCAATTAATGTAAGTGGTCCACCGCTCAACCA 3240  
Qy |||||  
Db 3346 AACGGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGCAGACTCTCTGCCATGTGAGA 3405  
Qy |||||  
Db 3241 AACGGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGCAGACTCTCTGCCATGTGAGA 3300  
Qy |||||  
Db 3406 CCACCTCTGTTACATATGAGAGAGCATATATTTTGAATAATCTGGAATAATACACTGAT 3465  
Qy |||||  
Db 3301 CCACCTCTGTTACATATGAGAGAGCATATATTTTGAATAATCTGGAATAATACACTGAT 3360  
Qy |||||  
Db 3466 TATATATTAATAATTAATCTCCATCAACAGAAAGGGATTTCTGATACCTATATCTGCCAG 3525  
Qy |||||  
Db 3361 TATATATTAATAATTAATCTCCATCAACAGAAAGGGATTTCTGATACCTATATCTGCCAG 3420  
Qy |||||  
Db 3526 CTATACATCAAGACTGAAAGAGATGTCAGAGAACTTCCACCAATATCAACCTTTTAA 3585  
Qy |||||  
Db 3421 CTATACATCAAGACTGAAAGAGATGTCAGAGAACTTCCACCAATATCAACCTTTTAA 3480  
Qy |||||  
Db 3586 AACCTTTCCTCTACCTCAGTCTCTTATCATGCGATCCCCAGTAAAGCCAAATGTCGA 3645  
Qy |||||

Db 3481 AACCTTTCCTCTACCTCAGTCTCTTATCATGGATCCCCAGTAAAGCCAAATGTCGA 3540  
Qy |||||  
Db 3646 ATAATAAGTTATGATTTAACTTTTCAAGGACCAAAATGAAAATTTATTTCTTTCACTACTCT 3705  
Qy |||||  
Db 3541 ATAATAAGTTATGATTTAACTTTTCAAGGACCAAAATGAAAATTTATTTCTTTCACTACTCT 3600  
Qy |||||  
Db 3706 GATAATTACATATATTTGGAAGAGCTTTTCAACATTTTACATTTATATATATAGCTTTTGTGCTCC 3765  
Qy |||||  
Db 3601 GATAATTACATATATTTGGAAGAGCTTTTCAACATTTTACATTTATATATATAGCTTTTGTGCTCC 3660  
Qy |||||  
Db 3766 GCAAGAACTAGAAAAAGGACTTGTCTCTCCAGTATTCTTTCTTTTACACAGATGAGTCA 3825  
Qy |||||  
Db 3661 GCAAGAACTAGAAAAAGGACTTGTCTCTCCAGTATTCTTTTCTTTTACACAGATGAGTCA 3720  
Qy |||||  
Db 3826 GTGCCGTTAGCACCTCCACAAAAATTTGACTTTTAACTCAACTGTACTTTAGATTTGATGG 3885  
Qy |||||  
Db 3721 GTGCCGTTAGCACCTCCACAAAAATTTGACTTTTAACTCAACTGTACTTTAGATTTGATGG 3780  
Qy |||||  
Db 3886 CTGAATGAGGCCCAAGTCTCTTCCAGTGGTATTGTTAAAGTATATATAGTTTAAATTT 3945  
Qy |||||  
Db 3781 CTGAATGAGGCCCAAGTCTCTTCCAGTGGTATTGTTAAAGTATATATAGTTTAAATTT 3840  
Qy |||||  
Db 3946 CATGAACATGAACCTGACACTATATATTAAGAATATATCAGGATTTAAACCTGAAGCC 4005  
Qy |||||  
Db 3841 CATGAACATGAACCTGACACTATATATTAAGAATATATCAGGATTTAAACCTGAAGCC 3900  
Qy |||||  
Db 4006 AAACCTTTGTTGGACTGGAACCAAGTCAAGCCTACTCTATCCGTGTATCTCGTTTCAACAA 4065  
Qy |||||  
Db 3901 AAACCTTTGTTGGACTGGAACCAAGTCAAGCCTACTCTATCCGTGTATCTCGTTTCAACAA 3960  
Qy |||||  
Db 4066 GTTGGAAATGGCAATCAATTTTAGTAAATAGTAAATTTCAACCCCAAGATCAGTTCCA 4125  
Qy |||||  
Db 3961 GTTGGAAATGGCAATCAATTTTAGTAAATAGTAAATTTCAACCCCAAGATCAGTTCCA 4020  
Qy |||||  
Db 4126 GATGCTGTCAGAAATATGAGTGCATAGCAACTAGCTGCGAGTCACTTTTAGTGAATGG 4185  
Qy |||||  
Db 4021 GATGCTGTCAGAAATATGAGTGCATAGCAACTAGCTGCGAGTCACTTTTAGTGAATGG 4080  
Qy |||||  
Db 4186 GATCAACCCAAAAAGGCAATGGAATAATAAGCAGTATATGTTAAAGGATTTGAAAGGAAT 4245  
Qy |||||  
Db 4081 GATCAACCCAAAAAGGCAATGGAATAATAAGCAGTATATGTTAAAGGATTTGAAAGGAAT 4140  
Qy |||||  
Db 4246 TCTACAAAAGTTTCTCCCAAGATCAATGATGATCACTTTTCAATAAGCTTTCTGCAATACC 4305  
Qy |||||  
Db 4141 TCTACAAAAGTTTCTCCCAAGATCAATGATGATCACTTTTCAATAAGCTTTCTGCAATACC 4200  
Qy |||||  
Db 4306 TCAATATGCTTTTAAAGTAAAGGCTTCAACCTCAGCTGGTGAAGGTCATGAAGCAGATGC 4365  
Qy |||||  
Db 4201 TCAATATGCTTTTAAAGTAAAGGCTTCAACCTCAGCTGGTGAAGGTCATGAAGCAGATGC 4260  
Qy |||||  
Db 4366 CATGTCAGCACACTACTCTGAAAACAGTTCCAGTGTTCACCAAAATATGCTTTTCTGAT 4425  
Qy |||||  
Db 4261 CATGTCAGCACACTACTCTGAAAACAGTTCCAGTGTTCACCAAAATATGCTTTTCTGAT 4320  
Qy |||||  
Db 4426 GTTCAGTCAACTAGTGCAACATTTGATGATGAAGTCAAGTCACTATCTTGGCTACTTTT 4485  
Qy |||||  
Db 4321 GTTCAGTCAACTAGTGCAACATTTGATGATGAAGTCAAGTCACTATCTTGGCTACTTTT 4380  
Qy |||||  
Db 4486 CAAATTTACAAATTTACCACTCAACTCTGCTGCTCAAAATGCAAGAAATGGGAATCCGA 4545  
Qy |||||  
Db 4381 CAAATTTACAAATTTACCACTCAACTCTGCTGCTCAAAATGCAAGAAATGGGAATCCGA 4440  
Qy |||||  
Db 4546 GAATGTTGAATATCAAAAAATTTCAATACCTCTATGAAGCTCACTTAACTGAAGAGACA 4605  
Qy |||||  
Db 4441 GAATGTTGAATATCAAAAAATTTCAATACCTCTATGAAGCTCACTTAACTGAAGAGACA 4500  
Qy |||||  
Db 4606 GTATATGGAATTAAGAAATTTAGATGATATAGATTTCAAGTGGCTGCCAGCAATGCT 4665  
Qy |||||  
Db 4501 GTATATGGAATTAAGAAATTTAGATGATATAGATTTCAAGTGGCTGCCAGCAATGCT 4560  
Qy |||||  
Db 4666 GGCATGGAATGCTTTCAAACTGGATTTCTACAAAACTTCGCTGGCTCCAGATGGT 4725  
Qy |||||  
Db 4561 GGCATGGAATGCTTTCAAACTGGATTTCTACAAAACTTCGCTGGCTCCAGATGGT 4620  
Qy |||||

|    |      |  |      |      |   |      |
|----|------|--|------|------|---|------|
| QY | 4726 | CCTCCTGAAATGTTTCATGCTAGTAGCAACATCACCTTTTGTAGCATCAGCATAGCTGGAGT | 4785 | 5806 | GAAAGAACGCTAGAGATCATTTCTTCGTCACTTTGTGTATCTTTTCAATAAATTTCTCTT  | 5865 |
| DB | 4621 | CCTCCTGAAATGTTTCATGCTAGTAGCAACATCACCTTTTGTAGCATCAGCATAGCTGGAGT | 4680 | 5701 | GAAAGAACGCTAGAGATCATTTCTTCGTCACTTTGTGTATCTTTTCAATAAATTTCTCTT  | 5760 |
| QY | 4786 | GAACTGCTGTCATTACTACGCAACCATGCTTATCTGATTTGATGTCGATCGGTAGATAAT   | 4845 | 5866 | GGAACAGCTATTTTGTGCAATTTGCAAGAAATTCGACAGAGCAAGAAAGGTGGCAGATAC  | 5925 |
| DB | 4681 | GAACTGCTGTCATTACTACGCAACCATGCTTATCTGATTTGATGTCGATCGGTAGATAAT   | 4740 | 5761 | GGAACAGCTATTTTGTGCAATTTGCAAGAAATTCGACAGAGCAAGAAAGGTGGCAGATAC  | 5820 |
| QY | 4846 | GATGAATTTAATATATATCTTCATCAAGTCAATGAAGAAATTAAGAACCATAGAAATTA    | 4905 | 5926 | TCCTCTCAGGATGCAAGAAATTTATTCACACTAAATTTGAAGCTGCATCAGCTCATCAGTG | 5985 |
| DB | 4741 | GATGAATTTAATATATATCTTCATCAAGTCAATGAAGAAATTAAGAACCATAGAAATTA    | 4800 | 5821 | TCCTCTCAGGATGCAAGAAATTTATTCACACTAAATTTGAAGCTGCATCAGCTCATCAGTG | 5880 |
| QY | 4906 | GATTTAGAAATTTCAAGGTATCTGTAGTGATCACTGCATTTTACTGCGAACATTA        | 4965 | 5986 | CGAGACCTCGAACTGAGGACGAGAGATTAACGCG-----                       | 6020 |
| DB | 4801 | GATTTAGAAATTTCAAGGTATCTGTAGTGATCACTGCATTTTACTGCGAACATTA        | 4860 | 5881 | CGAGACCTCGAACTGAGGACGAGAGATTAACGCGGTTACTTAGTTATAGAAATCCATC    | 5940 |
| QY | 4966 | GCTGCATATGTAGAGGGAAGTCAAGTCTGAAATGATTTGTACTACTTTTAGAATCAGCC    | 5025 | 6021 | --GCCAATAAGCAAGAAATCTTCTGTCGAAACATGTTGAAGAGCTTTTGCAAAACAAAC   | 6078 |
| DB | 4861 | GCTGCATATGTAGAGGGAAGTCAAGTCTGAAATGATTTGTACTACTTTTAGAATCAGCC    | 4920 | 5941 | AAGCCAATAAGCAAGAAATCTTCTGTCGAAACATGTTGAAGAGCTTTTGCAAAACAAAC   | 6000 |
| QY | 5026 | CCAAAGGACCCCTAAACATGATTTTCAAGATACCAAGTGAAGTTACAAATTT           | 5085 | 6079 | CTAAGATTTCAAGAGAAATTTTCCGAATTTACCAAAATTTCTTCAGATCTTTCTTCAACT  | 6138 |
| DB | 4921 | CCAAAGGACCCCTAAACATGATTTTCAAGATACCAAGTGAAGTTACAAATTT           | 4980 | 6001 | CTAAGATTTCAAGAGAAATTTTCCGAATTTACCAAAATTTCTTCAGATCTTTCTTCAACT  | 6060 |
| QY | 5086 | CAATTAACGTTCTCTCTCTCAACCTTAATGGAATATCCAAGTATATCAAGCTCTG        | 5145 | 6139 | GATGCTGATCTGCTTTGGAATAGAGCAAAACCCGTTTCCCAAAACATATAATAT        | 6198 |
| DB | 4981 | CAATTAACGTTCTCTCTCTCAACCTTAATGGAATATCCAAGTATATCAAGCTCTG        | 5040 | 6061 | GATGCTGATCTGCTTTGGAATAGAGCAAAACCCGTTTCCCAAAACATATAATAT        | 6117 |
| QY | 5146 | GTTTACCGAGAGATGATCTACTGCTGTCAGATTTCAACCTCAGTATTTATACAGAA       | 5205 | 6199 | AATAATAACAGAGTAAAGCTGATGACGCTAGTGTTCAGGTTTCAGGTTTCGATTTATTAAT | 6258 |
| DB | 5041 | GTTTACCGAGAGATGATCTACTGCTGTCAGATTTCAACCTCAGTATTTATACAGAA       | 5100 | 6118 | AATAATAACAGAGTAAAGCTGATGACGCTAGTGTTCAGGTTTCGATTTATTAAT        | 6177 |
| QY | 5206 | ACCAACACATTTGTCATTTGCAATGTAGAGGACTAAAGGTTGGAATATCAATATC        | 5265 | 6259 | GCCAGCTATTTCTGCTTTATTTATGTCCAAAATGATTTATTTGCTACTCAAGGTCACAT   | 6318 |
| DB | 5101 | ACCAACACATTTGTCATTTGCAATGTAGAGGACTAAAGGTTGGAATATCAATATC        | 5160 | 6178 | GCCAGCTATTTCTGCTTTATTTATGTCCAAAATGATTTATTTGCTACTCAAGGTCACAT   | 6237 |
| QY | 5266 | AGTGTTTACGAGTCAATAGTCTGCTGAGTCCAAAGTTCCGATGAGAAATACCATG        | 5325 | 6319 | CCAGGAAACAGTTGGAGATTTTGGAGAAATGTTGCGAAACAGGCGCAAAACATTTAGTA   | 6378 |
| DB | 5161 | AGTGTTTACGAGTCAATAGTCTGCTGAGTCCAAAGTTCCGATGAGAAATACCATG        | 5220 | 6238 | CCAGGAAACAGTTGGAGATTTTGGAGAAATGTTGCGAAACAGGCGCAAAACATTTAGTA   | 6297 |
| QY | 5326 | GATATCAAAAGCTCCAGCAGCAACAAACCAACCCCTATTTATGATGCCACAGGA         | 5385 | 6379 | ATGCTAAACACAGTGTGTTGAAAGGACGAGATCAGATGCAATCAGTATTTGGCAGAGGAC  | 6438 |
| DB | 5221 | GATATCAAAAGCTCCAGCAGCAACAAACCAACCCCTATTTATGATGCCACAGGA         | 5280 | 6298 | ATGCTAAACACAGTGTGTTGAAAGGACGAGATCAGATGCAATCAGTATTTGGCAGAGGAC  | 6357 |
| QY | 5386 | AAACTGCTTGTGATTTCAACAACATTAACAATCAGAAATGCCAATATGTTACTACAGTAT   | 5445 | 6439 | ACACAGCCAGTGTACTGCTTTGAGATATAGTATTTACAAAGCTAAATGAGGATGTTCAA   | 6498 |
| DB | 5281 | AAACTGCTTGTGATTTCAACAACATTAACAATCAGAAATGCCAATATGTTACTACAGTAT   | 5340 | 6358 | AAACAGCCAGTGTACTGCTTTGAGATATAGTATTTACAAAGCTAAATGAGGATGTTCAA   | 6417 |
| QY | 5446 | GATCATGGACCAATAAATAATGTACAAGTCTTGCGACAGAAACAGGAGCTCAGCATGAT    | 5505 | 6499 | ATAGATTTGGACTATCAGGGATCTGAAATTTGAAAGGCAATGGGATTTGCAATGCTGTTCA | 6558 |
| DB | 5341 | GATCATGGACCAATAAATAATGTACAAGTCTTGCGACAGAAACAGGAGCTCAGCATGAT    | 5400 | 6418 | ATAGATTTGGACTATCAGGGATCTGAAATTTGAAAGGCAATGGGATTTGCAATGCTGTTCA | 6477 |
| QY | 5506 | GGAATGTAAAGTGGTATGATGATATTTTAAATAAGCAAGGCGCATATTTTACAAAT       | 5565 | 6559 | CAGTGTAACTTTTACTGCTGCGCAGAGCATGCGGTTCTTGAGAAACAGCGCCCTCTAAT   | 6618 |
| DB | 5401 | GGAATGTAAAGTGGTATGATGATATTTTAAATAAGCAAGGCGCATATTTTACAAAT       | 5460 | 6478 | CAGTGTAACTTTTACTGCTGCGCAGAGCATGCGGTTCTTGAGAAACAGCGCCCTCTAAT   | 6537 |
| QY | 5566 | GAAAGCTTTCTAAACCTCATGTAAGAAGGAAACAAAGTTTATGTCGCAATGAAGAA       | 5625 | 6619 | CACTTTGTGAAGTGTGTTGAGACAGGCGCAGATGACACCAACCTATGATTTGTTTAC     | 6678 |
| DB | 5461 | GAAAGCTTTCTAAACCTCATGTAAGAAGGAAACAAAGTTTATGTCGCAATGAAGAA       | 5520 | 6538 | CACTTTGTGAAGTGTGTTGAGACAGGCGCAGATGACACCAACCTATGATTTGTTTAC     | 6597 |
| QY | 5626 | ATCTACATCATAGTCTGATATGATGATGATGATGATGATGATGATGATGATGATGATG     | 5685 | 6679 | TCAGTGTGAGTGTGGAAGAACTGGAGTTTTTATTTGCTCTGAGCCATTTTAAACAAACAT  | 6738 |
| DB | 5521 | ATCTACATCATAGTCTGATATGATGATGATGATGATGATGATGATGATGATGATGATG     | 5580 | 6598 | TGCAGTGTGAGTGTGGAAGAACTGGAGTTTTTATTTGCTCTGAGCCATTTTAAACAAACAT | 6657 |
| QY | 5686 | AATGACCACTGAAACCAAAAGCAATATTTTAAATTTTAGAGCTACAAATATTATG        | 5745 | 6739 | ATAAATGACCAATGATTTTGTGATATATATGATGATGATGATGATGATGATGATGATG    | 6798 |
| DB | 5581 | AATGACCACTGAAACCAAAAGCAATATTTTAAATTTTAGAGCTACAAATATTATG        | 5640 | 6658 | ATAAATGACCAATGATTTTGTGATATATATGATGATGATGATGATGATGATGATGATG    | 6717 |
| QY | 5746 | GGAACATTTTACTGATCTGATTTTCTGACCCCTGTTTAAAGCTTTAGGGAAGGACTTTC    | 5805 | 6799 | ATGTCATGCTGCAAGTCTGCAAGTATATCTTTTACACAGTGCATTTCTGATCTC        | 6858 |
| DB | 5641 | GGAACATTTTACTGATCTGATTTTCTGACCCCTGTTTAAAGCTTTAGGGAAGGACTTTC    | 5700 | 6718 | ATGTCATGCTGCAAGTCTGCAAGTATATCTTTTACACAGTGCATTTCTGATCTC        | 6777 |
|    |      |  |      | 6859 | TTATCAAAATAGGGAAGTAAATCAGCCCATCTGTTTTGTTTAACTATTTACAGACTTTCAG | 6918 |



Db 6778 TTATCAATAGGAAGTAATCAGCCATCTGTTTGTAACTATTGAGCACTTCAGAAG 6837  
 Qy 6919 ATGAGCTCTTTGGAGCGCATGGAAGTGATGTTGAGCTTGAATGGGAAGAAACCACTATG 6978  
 Db 6838 ATGAGCTCTTTGGAGCGCATGGAAGTGATGTTGAGCTTGAATGGGAAGAAACCACTATG 6897

## RESULT 3

ABA00717

ID ABA00717 standard; cDNA; 7225 BP.

XX ABA00717;

XX 04-MAR-2003 (first entry)

XX Human KPP-2 cDNA, Incyte ID No. 7480588CB1.

XX Gene; kinase; phosphatase; KPP; cell proliferation; arteriosclerosis;  
 KW atherosclerosis; cirrhosis; hepatitis; reproduction; infertility;  
 KW paroxysmal nocturnal haemoglobinuria; polycythaemia vera; psoriasis;  
 KW primary thrombocytopenia; cancer; development; renal tubular acidosis;  
 KW anaemia; mental retardation; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; epilepsy; Sjogren's syndrome; uveitis; asthma;  
 KW menstrual cycle; autoimmune; inflammation; rheumatoid arthritis; AIDS;  
 KW autoimmune thyroiditis; contact dermatitis; Crohn's disease; allergy;  
 KW diabetes mellitus; glomerulonephritis; Goodpasture's syndrome; gout;  
 KW Graves' disease; Hashimoto's thyroiditis; irritable bowel syndrome;  
 KW multiple sclerosis; osteoarthritis; osteoporosis; pancreatitis;  
 KW Reiter's syndrome; rat; glomerular mesangial cell receptor protein;  
 KW tyrosine phosphatase precursor; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 1..6900

XX FT /\*tag= a

XX FT /product= "KPP-2"

XX WO200290530-A2.

XX 14-NOV-2002.

XX 16-JAN-2002; 2002WO-US001369.

XX 18-JAN-2001; 2001US-0263083P.

XX 23-FEB-2001; 2001US-0271117P.

XX 23-FEB-2001; 2001US-0271205P.

XX 16-MAR-2001; 2001US-0276859P.

XX 23-MAR-2001; 2001US-0278504P.

XX 23-MAR-2001; 2001US-0278522P.

XX 29-MAR-2001; 2001US-0280266P.

XX 29-MAR-2001; 2001US-0280510P.

XX (INCY-) INCYTE GENOMICS INC.

XX Lee EA, Wallia NK, Baughn MR, Ison CH, Gururajan R, Arvizu C;

XX PI Yao MG, Jackson JL, Tang TY, Yue H, Tran B, Ding L, Lu DAM;

XX PI Lal PG, Warren BA;

XX WPI; 2003-111972/10.

XX P-PSDB; AAG79724.

XX New human kinases and phosphatases and polynucleotides, useful for  
 PT diagnosing, treating or preventing autoimmune or inflammatory disorders  
 PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,  
 PT cancer or hepatitis.

XX Claim 5; Page 138-39; 144pp; English.

XX The sequences given in ABA00716-23 encode human kinases and phosphatases  
 CC (KPP). The KPP polypeptides, polynucleotides, and agonists and  
 CC antagonists to them, are useful for diagnosing, treating or preventing

CC disorders associated with aberrant expression of KPP, particularly cell  
 CC proliferative disorders (e.g. arteriosclerosis, atherosclerosis,  
 CC cirrhosis, hepatitis, paroxysmal nocturnal haemoglobinuria, polycythaemia  
 CC vera, psoriasis, primary thrombocytopenia or cancer), developmental  
 CC disorders (e.g. renal tubular acidosis, anaemia or mental retardation),  
 CC neurological disorders (e.g. Alzheimer's disease, Parkinson's disease or  
 CC epilepsy), reproductive disorders (e.g. infertility or a disruption in  
 CC the menstrual cycle), or autoimmune/inflammatory disorders (e.g. AIDS,  
 CC allergies, asthma, autoimmune thyroiditis, contact dermatitis, Crohn's  
 CC disease, diabetes mellitus, glomerulonephritis, Goodpasture's syndrome,  
 CC gout, Graves' disease, Hashimoto's thyroiditis, irritable bowel syndrome,  
 CC multiple sclerosis, osteoarthritis, osteoporosis, pancreatitis, Reiter's  
 CC syndrome, rheumatoid arthritis, Sjogren's syndrome, uveitis), or viral,  
 CC bacterial, fungal, parasitic, protozoan or helminthic infections. They  
 CC are also useful in the assessment of the effects of exogenous compounds  
 CC on the expression of nucleic acid and amino acid sequences of proteins  
 CC associated with KPP. The protein encoded by this sequence is homologous  
 CC to rat glomerular mesangial cell receptor protein tyrosine phosphatase  
 CC precursor

XX SQ Sequence 7225 BP; 2434 A; 1417 C; 1352 G; 2022 T; 0 U; 0 Other;

Query Match 95.7%; Score 6799.8; DB 8; Length 7225;

Best Local Similarity 98.7%; Pred. No. 0;

Matches 6962; Conservative 0; Mismatches 17; Indels 72; Gaps 8;

Qy 106 ATGAGTTTCTTATCATCTTTCTTTTCTTTTATTTGGGACTTTCAGAGACACAGGTTGAT 165

Db 1 ATGGATTTTCTTATCATCTTTCTTTTCTTTTATTTGGGACTTTCAGAGACACAGGTTGAT 60

Qy 166 GTTTCGAATGTCGTTCTCTGTTACTAGGTACGATATTAACCATCTCTTCAATTTCTACAACA 225

Db 61 GTTTCGAATGTCGTTCTCTGTTACTAGGTACGATATTAACCATCTCTTCAATTTCTACAACA 120

Qy 226 TACACCTCACCTGTTACTAGATAGTACACCAAAATGTAACAAACACAGGCTCCAGTC 285

Db 121 TACACCTCACCTGTTACTAGATAGT-----GAACAGAGGCTCCAGTC 165

Qy 286 TTCTAGCCGGGAAAGAGTCGGATCTGCTGGGATTTCTTCTTTGGAAATACACACCT 345

Db 166 TTCTAGCCGGGAAAGAGTCGGATCTGCTGGGATTTCTTCTTTGGAAATACACACCT 225

Qy 346 AATCCAAATGGAAGGATATATCTTACATGTCAATATAAGGAAGTTTGTCCGTGGATG 405

Db 226 AATCCAAATGGAAGGATATATCTTACATGTCAATATAAGGAAGTTTGTCCGTGGATG 285

Qy 406 CAAACAGTATATACCAAGTCAGNTCAAGCCAGACAGTCTGGAAGTTCTTCTTACTAAT 465

Db 286 CAAACAGTATATACCAAGTCAGNTCAAGCCAGACAGTCTGGAAGTTCTTCTTACTAAT 345

Qy 466 CTTAATCTCGGAACAAACATATGAAATTAAGGTTGCTGTAACAGTCTGGCATTTGGA 525

Db 346 CTTAATCTCGGAACAAACATATGAAATTAAGGTTGCTGTAACAGTCTGGCATTTGGA 405

Qy 526 GTGTTTAGTGATCCATTTCTTCTTCCAACTGCAGAAAGTCTCCAGGAAAAGTGGTGAAT 585

Db 406 GTGTTTAGTGATCCATTTCTTCTTCCAACTGCAGAAAGTCTCCAGGAAAAGTGGTGAAT 465

Qy 586 CTCACAGTTGAGCCCTACAAAGCTTCAGCAGTTAGCTGATTTGGTATTTACCTCGGCAA 645

Db 466 CTCACAGTTGAGCCCTACAAAGCTTCAGCAGTTAGCTGATTTGGTATTTACCTCGGCAA 525

Qy 646 CCAAAATGGCAAAATTACCAGCTTCAAGATTAGTGTCACAAATGTCAGGAAAGTGGATAGTA 705

Db 526 CCAAAATGGCAAAATTACCAGCTTCAAGATTAGTGTCACAAATGTCAGGAAAGTGGATAGTA 585

Qy 706 GTGAAGATGTCCTCAATCAGAGTAGAGGACATTTTGACTGGGAAATTTGCCAGAAATGCAAT 765

Db 586 GTGAAGATGTCCTCAATCAGAGTAGAGGACATTTTGACTGGGAAATTTGCCAGAAATGCAAT 645

Qy 766 GAGAATAGTGAATCTTTTATTTAGGAGTACAGCAGCCCTTCTCCAAACCTTGTGAGATT 825

Db 646 GAGAATAGTGAATCTTTTATTTAGGAGTACAGCAGCCCTTCTCCAAACCTTGTGAGATT 705

|    |      |  |      |    |      |   |      |
|----|------|--|------|----|------|---|------|
| QY | 826  | ACACCTCCATCGGTACCAACATTCATCAAGCACGTTGACACAGAAATGAGATCAGCTCT    | 885  | QY | 1900 | ATAAATCACTATACGATTTATGCAATGGAATTTGGATACAAACAGAGCATTTCCAGATAACT  | 1959 |
| DB | 706  | ACACCTCCATCGGTACCAACATTCATCAAGCACGTTGACACAGAAATGAGATCAGCTCT    | 765  | DB | 1783 | ATRAATCACTATACGATTTATGCAATGGAATTTGGATACAAACAGAGCATTTCCAGATAACT  | 1842 |
| QY | 886  | GTGTGAAAGAGCCTATCAGTTTTGTAGTGACACACTTGAGACCTTTATACAAATATCTT    | 945  | QY | 1960 | ACCATAGATAACAGCTTTCTCATACAG-----GGTTAAAGAAATACACAAA             | 2007 |
| DB | 766  | GTGTGAAAGAGCCTATCAGTTTTGTAGTGACACACTTGAGACCTTTATACAAATATCTT    | 825  | DB | 1843 | ACCATAGATAACAGCTTTCTCATACAGGTAGAAAAAATGTTTAAAGAAATACACAAA       | 1902 |
| QY | 946  | TTTGAAGTTTACGCTCTCAACTGAAGCAGGTTTATTTGATAGTACGATTTGTCAAGACA    | 1005 | QY | 2008 | TACAAAATGAGAGTGGCAGCCTCAACCCACGATGAGAAAGTCTTTGTCTGAGAAAT        | 2067 |
| DB | 826  | TTTGAAGTTTACGCTCTCAACTGAAGCAGGTTTATTTGATAGTACGATTTGTCAAGACA    | 885  | DB | 1903 | TACAAAATGAGAGTGGCAGCCTCAACCCACGTTGGAGAAAGTCTTTGTCTGAGAAAT       | 1962 |
| QY | 1006 | CCAGATCAGTCCGTAAGGACCAACCAAACTGCTAAACAGGCAACATCAAGGAAG         | 1065 | QY | 2068 | GACATCTTTGTGAGAACTTCAGAGATGAACCGGAATCATCACCTCAAGATGTCGAAGTA     | 2127 |
| DB | 886  | CCAGATCAGTCCGTAAGGACCAACCAAACTGCTAAACAGGCAACATCAAGGAAG         | 945  | DB | 1963 | GACATCTTTGTGAGAACTTCAGAGATGAACCGGAATCATCACCTCAAGATGTCGAAGTA     | 2022 |
| QY | 1066 | TCCTTTTCAATTTATGGGACCCACCAACTATAGTAAACAGGGAAATTTATAGTTATAGATT  | 1125 | QY | 2128 | ATTGATGTTTACCGCAGATGAATAAGGTTGAAAGTGGTCAACCCGAAAGCCCAATGGG      | 2187 |
| DB | 946  | TCCTTTTCAATTTATGGGACCCACCAACTATAGTAAACAGGGAAATTTATAGTTATAGATT  | 1005 | DB | 2023 | ATTGATGTTTACCGCAGATGAATAAGGTTGAAAGTGGTCAACCCGAAAGCCCAATGGG      | 2082 |
| QY | 1126 | GAATATATGGACCAT-----CAGGTGCGATTTTGGATAACAGCAACAAAGACCTCAAG     | 1179 | QY | 2188 | ATCATTTATGCTTATGAAAGTCTATATAAATAATATAGATCTTTATATATGAGAACACA     | 2247 |
| DB | 1006 | GAATATATGGACCAT-----CAGGTGCGATTTTGGATAACAGCAACAAAGACCTCAAG     | 1065 | DB | 2083 | ATCATTTATGCTTATGAAAGTCTATATAAATAATATAGATCTTTATATATGAGAACACA     | 2142 |
| QY | 1180 | TTTGCAATTCACCTAACCATTTTCAATGTATGATGCTTATATGCGGCTGAACCC         | 1239 | QY | 2248 | TCACAAACAGACATAATATTAAAGGAACCTTAAGACCTCAACCCCTCTATATACTTTCTGTA  | 2307 |
| DB | 1066 | TTTGCAATTCACCTAACCATTTTCAATGTATGATGCTTATATGCGGCTGAACCC         | 1125 | DB | 2143 | TCACAAACAGACATAATATTAAAGGAACCTTAAGACCTCAACCCCTCTATATACTTTCTGTA  | 2202 |
| QY | 1240 | AGTGAGGACCTGGGCCCCAAGTCAAAATATTTCAAGTATTCATCTCCACAGATGTTCCAGGG | 1299 | QY | 2308 | AGGCTCTTACACAGATTTGGTCAATGCGCAATCAGGATATCTTTTACTCTCTGTGAAGACT   | 2367 |
| DB | 1126 | AGTGAGGACCTGGGCCCCAAGTCAAAATATTTCAAGTATTCATCTCCACAGATGTTCCAGGG | 1185 | DB | 2203 | AGGCTCTTACACAGATTTGGTCAATGCGCAATCAGGATATCTTTTACTCTCTGTGAAGACT   | 2262 |
| QY | 1300 | GCAGTGTGTTGATTACAACTTGAGAGGTAGAAATCCACCAAGTAAGAAATTTCTGGAAG    | 1359 | QY | 2368 | TCGGAGACTGTGCTGATAGTGCAACAGAAATATACCTTACAAATAATTTCTTCTGCA       | 2427 |
| DB | 1186 | GCAGTGTGTTGATTACAACTTGAGAGGTAGAAATCCACCAAGTAAGAAATTTCTGGAAG    | 1245 | DB | 2263 | TCGGAGACTGTGCTGATAGTGCAACAGAAATATACCTTACAAATAATTTCTTCTGCA       | 2322 |
| QY | 1360 | AAACACGACCAACCAATGGAATTTAACCAATACCGAGTGAAGTGTCTCCAGAG          | 1419 | QY | 2428 | GAGATTGAGCTATCATTTCTTCCCAAGTAGTCCCAATGGAATCATATAAATAATATACA     | 2487 |
| DB | 1246 | AAACACGACCAACCAATGGAATTTAACCAATACCGAGTGAAGTGTCTCCAGAG          | 1305 | DB | 2323 | GAGATTGAGCTATCATTTCTTCCCAAGTAGTCCCAATGGAATCATATAAATAATATACA     | 2382 |
| QY | 1420 | ACAGGAATAATTTTGAAATACTTTGCTCACTGGAAATAATGAGTATATAATGACCCC      | 1479 | QY | 2488 | ATTATCTCAAGAGAAATGGAATAAGAGAAAGAACTATATAAATAAATCTTTTAAAC        | 2547 |
| DB | 1306 | ACAGGAATAATTTTGAAATACTTTGCTCACTGGAAATAATGAGTATATAATGACCCC      | 1362 | DB | 2383 | ATTATCTCAAGAGAAATGGAATAAGAGAAAGAACTATATAAATAAATCTTTTAAAC        | 2442 |
| QY | 1480 | ATGGCTCCAGAAATTTGGAACATAGTAGAGCCCAATCGTAGGATATATAGAGGTTGAGCA   | 1539 | QY | 2548 | CAAAACATTAAGTACTGAGAAATATACCCCAATATATCATTTGAGGTGCTCTGTAGTACA    | 2607 |
| DB | 1363 | ATGGCTCCAGAAATTTGGAACATAGTAGAGCCCAATCGTAGGATATATAGAGGTTGAGCA   | 1422 | DB | 2443 | CAAAACATTAAGTACTGAGAAATATACCCCAATATATCATTTGAGGTGCTCTGTAGTACA    | 2502 |
| QY | 1540 | GAGATGCTGCTGACCTTCACTCACTGCTACATTTATATATACAGCCATCCAGATAA       | 1599 | QY | 2608 | CTGAAAGGTGAAGGAGTTCGGAGTGTCTCCATAAGTATATACTGACCGGAGGAAGATGCTCT  | 2667 |
| DB | 1423 | GAGATGCTGCTGACCTTCACTCACTGCTACATTTATATATACAGCCATCCAGATAA       | 1482 | DB | 2503 | CTCAAGGTGAAGGAGTTCGGAGTGTCTCCATAAGTATATACTGACCGGAGGAAGATGCTCT   | 2562 |
| QY | 1600 | AACTTTCTGCAAGGAATAGAGCTGAAGACCAAGCTTCCACAGTGTGTAACATAAGGAAT    | 1659 | QY | 2668 | GATTCTCCCTTCAAGACTTCTCTGTAAACAGTGTCTGTGTGTGTGTGTGTGTGTGTGTGT    | 2727 |
| DB | 1483 | AACTTTCTGCAAGGAATAGAGCTGAAGACCAAGCTTCCACAGTGTGTAACATAAGGAAT    | 1542 | DB | 2563 | GATTCTCCCTTCAAGACTTCTCTGTAAACAGTGTCTGTGTGTGTGTGTGTGTGTGTGTGT    | 2622 |
| QY | 1660 | CAGTATATTAATGACATTTGAGCTGAACAGCTGTCTTATGTTATACAGGAGCTTGTACCT   | 1719 | QY | 2728 | TGGCAACCACTTGGAGCCAAATGGAATTTATCTTTATATACAGTGTGTGTGTGTGTGTGT    | 2784 |
| DB | 1543 | CAGTATATTAATGACATTTGAGCTGAACAGCTGTCTTATGTTATACAGGAGCTTGTACCT   | 1602 | DB | 2623 | TGGCAACCACTTGGAGCCAAATGGAATTTATCTTTATATACAGTGTGTGTGTGTGTGTGT    | 2682 |
| QY | 1720 | TTTCACTGAGCAGATGATGATGATCTGCTTTTCCACCATCATGCGGAGAGGACCAACACA   | 1779 | QY | 2785 | AATAGATCATTAATAAATAATTAATGTCTCTGAAACATCATTTGGAGTGTATCAGATTTG    | 2844 |
| DB | 1603 | TTTCACTGAGCAGATGATGATGATCTGCTTTTCCACCATCATGCGGAGAGGACCAACACA   | 1662 | DB | 2683 | AATAGATCATTAATAAATAATTAATGTCTCTGAAACATCATTTGGAGTGTATCAGATTTG    | 2742 |
| QY | 1780 | GTCTCAGTGTGAGGACCTGAGCAAGTCCCAAGCTCCATTAATAAATAATAAATAATAA     | 1839 | QY | 2845 | GATTATAATGTTGAATACAGTCTTTATGTAAACAGCTGTGTGTGTGTGTGTGTGTGTGTGT   | 2904 |
| DB | 1663 | GTCTCAGTGTGAGGACCTGAGCAAGTCCCAAGCTCCATTAATAAATAATAAATAATAA     | 1722 | DB | 2743 | GATTATAATGTTGAATACAGTCTTTATGTAAACAGCTGTGTGTGTGTGTGTGTGTGTGTGT   | 2802 |
| QY | 1840 | AATATTAGTCTTCTCATTTTGTGTTATATTTGGATCTTCCAGAAATATCCCAATGGAAA    | 1899 | QY | 2905 | ACAGGAGCAATATCTTAGCTTTTCAACACAGAGGAGCACCAGCATCTCTCCAAA          | 2964 |
| DB | 1723 | AATATTAGTCTTCTCATTTTGTGTTATATTTGGATCTTCCAGAAATATCCCAATGGAAA    | 1782 | DB | 2803 | ACAAGAGCAATATCTTAGCTTTTCAACACAGAGGAGG---ACCAAGCATCTCTCCAAA      | 2859 |
|    |      |  |      | QY | 2965 | GATGTTTATATGCAAACTCAGTCTTCTCATTAATAATTTCTTTTCTGTGAGACACCTCTCTCA | 3024 |

Db 2860 GATGTTTATATGCAACCTCAGTCTCTCATCAATAATCTTTCTGGACACCTCCTTCA 2919  
QY 3025 AAACCTAATGGGATATACAAATATTAATCTGTTTATACAGAAATATCTTCAGGTACTTTT 3084  
Db 2920 AAACCTAATGGGATATACAAATATTAATCTGTTTATACAGAAATATCTTCAGGTACTTTT 2979  
QY 3085 ATGCAGAAATTTTACATCCATGAATCAATCAATGAATTTGACAAATATGATGATCCACA 3144  
Db 2980 ATGCAGAAATTTTACATCCATGAATCAATCAATGAATTTGACAAATATGATGATCCACA 3039  
QY 3145 ATTATAGATAAATGACAAATATTAATCTGTTTATACAGAAATATCTTCAGGTACTTTT 3204  
Db 3040 ATTATAGATAAATGACAAATATTAATCTGTTTATACAGAAATATCTTCAGGTACTTTT 3099  
QY 3205 GTTGGAAATGGGAATAAAGCAGTGCATCAATGAAGTATACAGATCAAGACATACCT 3264  
Db 3100 GTTGGAAATGGGAATAAAGCAGTGCATCAATGAAGTATACAGATCAAGACATACCT 3159  
QY 3265 GAAGGTTTGTGGAAACCTGACTTACGAATCAATTTGCTCAACTGCAATGAAGTAAAGC 3324  
Db 3160 GAAGGTTTGTGGAAACCTGACTTACGAATCAATTTGCTCAACTGCAATGAAGTAAAGC 3219  
QY 3325 TGGGTCCACCGGCTCAACCAAGCTGCTAGTCTCTACTATGTTTCACTGATCTTACAG 3384  
Db 3220 TGGGTCCACCGGCTCAACCAAGCTGCTAGTCTCTACTATGTTTCACTGATCTTACAG 3279  
QY 3385 CAGACTCTCGCCATGTGAGACCACTCTGTTTACATATGAGAGAGACATATATTTGAT 3444  
Db 3280 CAGACTCTCGCCATGTGAGACCACTCTGTTTACATATGAGAGAGACATATATTTGAT 3339  
QY 3445 AATCTGGAATAATACACTGATATATATTAATAATTAATCTCATCAACAGAAAGGATTC 3504  
Db 3340 AATCTGGAATAATACACTGATATATATTAATAATTAATCTCATCAACAGAAAGGATTC 3399  
QY 3505 TCTGATACCTTATCTGCCAGCTATACATCAAGCTGAAGAGATGTCACAGAACTTCA 3564  
Db 3400 TCTGATACCTTATCTGCCAGCTATACATCAAGCTGAAGAGATGTCACAGAACTTCA 3459  
QY 3565 CCAATAATCAACACTTTTAAACCTTCTCTACCTCAGTCTCTTATCATGGGATCCC 3624  
Db 3460 CCAATAATCAACACTTTTAAACCTTCTCTACCTCAGTCTCTTATCATGGGATCCC 3519  
QY 3625 CCAGTAAAGCAAATGGTGCATTAATTAAGTTATGATTTAACTTTTACAGGACCAATGAA 3684  
Db 3520 CCAGTAAAGCAAATGGTGCATTAATTAAGTTATGATTTAACTTTTACAGGACCAATGAA 3579  
QY 3685 AATTATCTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3744  
Db 3580 AATTATCTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3639  
QY 3745 TTATATAGCTTTTGTGCGCGCAAGAACTAGAAAAAGGACTTGGTCTCTCCAGTATCTT 3804  
Db 3640 TTATATAGCTTTTGTGCGCGCAAGAACTAGAAAAAGGACTTGGTCTCTCCAGTATCTT 3699  
QY 3805 TTCTTTTACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3864  
Db 3700 TTCTTTTACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3759  
QY 3865 TGTACTTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGATGTT 3924  
Db 3760 TGTACTTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGATGTT 3819  
QY 3925 AAAGTATATAGTTTAAATTCATGAACATGAACATGAACATGAACATGAACATGAACATGA 3984  
Db 3820 AAAGTATATAGTTTAAATTCATGAACATGAACATGAACATGAACATGAACATGAACATGA 3879  
QY 3985 TCAGGATTTAAACTGAAGCAAACCTTGTGGACTGGAACCAAGTCAAGCACTTCTATC 4044  
Db 3880 TCAGGATTTAAACTGAAGCAAACCTTGTGGACTGGAACCAAGTCAAGCACTTCTATC 3939  
QY 4045 CGGTATCTCGGTTCACCAAGTTGGAATGGCAATCAATTTAGTAAATTTAGTAAATTTAGT 4104

Db 3940 CGGTATCTCGGTTCACCAAGTTGGAATGGCAATCAATTTAGTAAATTTAGTAAATTTAGT 3999  
QY 4105 ACAACCAAGAAATCAGTCTCCAGATGTCGTCAGAAATATGCAATGCAATGCAATGCAATG 4164  
Db 4000 ACAACCAAGAAATCAGTCTCCAGATGTCGTCAGAAATATGCAATGCAATGCAATGCAATG 4059  
QY 4165 CAGTCAATTTAGTAAATGGGATCCACCAAAAGGCAAAATGGAATTAATTAAGCAGTAT 4224  
Db 4060 CAGTCAATTTAGTAAATGGGATCCACCAAAAGGCAAAATGGAATTAATTAAGCAGTAT 4119  
QY 4225 ATGGTAAAGGAAATCTACAAAAGTTTCTCCCAAGATCAATGATCAATGATCAATGAT 4284  
Db 4120 ATGGTAAAGGAAATCTACAAAAGTTTCTCCCAAGATCAATGATCAATGATCAATGAT 4179  
QY 4285 ATAAAGCTTTCTGGCAATACCTCATATGTTTAAAGTAAAGCTTTCAACCTCAGCTGGT 4344  
Db 4180 ATAAAGCTTTCTGGCAATACCTCATATGTTTAAAGTAAAGCTTTCAACCTCAGCTGGT 4239  
QY 4345 GAAGGTGATGAAGACATGCTCCATGTCAGACACTACCTGAAACAGTTCAGTGTCTCC 4404  
Db 4240 GAAGGTGATGAAGACATGCTCCATGTCAGACACTACCTGAAACAGTTCAGTGTCTCC 4299  
QY 4405 ACAATATTTGCTTTTCTGATGTTTCACTCAATGTCAGTCAATGTCAGTCAATGTCAGT 4464  
Db 4300 ACAATATTTGCTTTTCTGATGTTTCACTCAATGTCAGTCAATGTCAGTCAATGTCAGT 4359  
QY 4465 GACACTATCTTGGCTACTTTTCAAAATTAACAAATTTACCACTCAACTGCTCAAAA 4524  
Db 4360 GACACTATCTTGGCTACTTTTCAAAATTAACAAATTTACCACTCAACTGCTCAAAA 4419  
QY 4525 TGCAAAGAAATGGGAATCCGAAAGATGTTGTAATTAATCAAAATTTCAATCACTCTATGAA 4584  
Db 4420 TGCAAAGAAATGGGAATCCGAAAGATGTTGTAATTAATCAAAATTTCAATCACTCTATGAA 4479  
QY 4585 GCTCACTTAACCTGAAGACAGTATATGATTAATGAATTAAGAAATTTAGTAAATTTAGT 4644  
Db 4480 GCTCACTTAACCTGAAGACAGTATATGATTAATGAATTAAGAAATTTAGTAAATTTAGT 4539  
QY 4645 GTGGCTGCCAGACCACTGCTGCTATGCAATGCTTCAAACTGGATTTCTACAAAACCT 4704  
Db 4540 GTGGCTGCCAGACCACTGCTGCTATGCAATGCTTCAAACTGGATTTCTACAAAACCT 4599  
QY 4705 GTGGCTGCCCTCCAGATGCTCTCTGAAATGTTTCAATGATGATGATGATGATGATGAT 4764  
Db 4600 GTGGCTGCCCTCCAGATGCTCTCTGAAATGTTTCAATGATGATGATGATGATGATGAT 4659  
QY 4765 AGCATCAGCATTAAGCTGAGTGAACCTGCTGCTATGCTGGAACCAATGATGATGATGAT 4824  
Db 4660 AGCATCAGCATTAAGCTGAGTGAACCTGCTGCTATGCTGGAACCAATGATGATGATGAT 4719  
QY 4825 GATGTCAAATCGGTAGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 4884  
Db 4720 GATGTCAAATCGGTAGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 4779  
QY 4885 AATAAAACATAGAAATTAAGATTTAGAAATTAATTAATTAATTAATTAATTAATTAATTA 4944  
Db 4780 AATAAAACATAGAAATTAAGATTTAGAAATTAATTAATTAATTAATTAATTAATTAATTA 4839  
QY 4945 GCATTTACTGGGAACATTAAGTCTGATATGATGATGATGATGATGATGATGATGATGAT 5004  
Db 4840 GCATTTACTGGGAACATTAAGTCTGATATGATGATGATGATGATGATGATGATGATGAT 4899  
QY 5005 GTTACTACTTTAGAAATCAGCCCCAAGGACCCACCTTAACAAATGATGATGATGATGATGAT 5064  
Db 4900 GTTACTACTTTAGAAATCAGCCCCAAGGACCCACCTTAACAAATGATGATGATGATGATGAT 4959  
QY 5065 CCAGATGAAGTTTACAAAATTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5124  
Db 4960 CCAGATGAAGTTTACAAAATTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5019  
QY 5125 ATCCAGATGATTAATCAAGCTCTGGTTTACCGAAGATGATGATGATGATGATGATGATGAT 5184  
Db 5020 ATCCAGATGATTAATCAAGCTCTGGTTTACCGAAGATGATGATGATGATGATGATGATGAT 5079



nootropic; antiparkinsonian; antiasthmatic; antiinfertility;  
cardiomyopathy; atherosclerosis; hypertension; cancer; obesity; diabetes;  
AIDS; multiple sclerosis; graft-versus-host disease; Alzheimer's disease;  
Parkinson's disease; asthma; fertility disorder; chromosome mapping;  
tissue typing; preventive medicine; pharmacogenomic; vaccine.

Homo sapiens.

WO2003102159-A2.

11-DEC-2003.

04-JUN-2003; 2003WO-US017573.

04-JUN-2002; 2002US-0385490P.

04-JUN-2002; 2002US-0385615P.

04-JUN-2002; 2002US-0385755P.

05-JUN-2002; 2002US-0386041P.

06-JUN-2002; 2002US-0386355P.

06-JUN-2002; 2002US-0386357P.

06-JUN-2002; 2002US-0386447P.

06-JUN-2002; 2002US-0386459P.

06-JUN-2002; 2002US-0386459P.

06-JUN-2002; 2002US-0386459P.

07-JUN-2002; 2002US-0386701P.

07-JUN-2002; 2002US-0386796P.

07-JUN-2002; 2002US-0386931P.

07-JUN-2002; 2002US-0387081P.

07-JUN-2002; 2002US-0387083P.

10-JUN-2002; 2002US-0387429P.

10-JUN-2002; 2002US-0387540P.

10-JUN-2002; 2002US-0387856P.

11-JUN-2002; 2002US-0387856P.

11-JUN-2002; 2002US-0387859P.

11-JUN-2002; 2002US-0387859P.

12-JUN-2002; 2002US-0387934P.

12-JUN-2002; 2002US-0387960P.

12-JUN-2002; 2002US-0388022P.

12-JUN-2002; 2002US-0388096P.

12-JUN-2002; 2002US-0388432P.

12-JUN-2002; 2002US-0388479P.

13-JUN-2002; 2002US-0389123P.

14-JUN-2002; 2002US-0389120P.

14-JUN-2002; 2002US-0389146P.

17-JUN-2002; 2002US-0389742P.

18-JUN-2002; 2002US-0389604P.

18-JUN-2002; 2002US-0389884P.

19-JUN-2002; 2002US-0390006P.

19-JUN-2002; 2002US-0390144P.

19-JUN-2002; 2002US-0390209P.

25-JUN-2002; 2002US-0391726P.

06-AUG-2002; 2002US-0401628P.

09-AUG-2002; 2002US-0402268P.

12-AUG-2002; 2002US-0402822P.

13-AUG-2002; 2002US-0403458P.

15-AUG-2002; 2002US-0403617P.

26-AUG-2002; 2002US-0403732P.

12-SEP-2002; 2002US-0406182P.

13-SEP-2002; 2002US-0410085P.

23-SEP-2002; 2002US-0412955P.

30-SEP-2002; 2002US-0415195P.

23-OCT-2002; 2002US-0420627P.

23-OCT-2002; 2002US-0420718P.

24-OCT-2002; 2002US-0420852P.

31-OCT-2002; 2002US-0422750P.

01-NOV-2002; 2002US-0423095P.

05-NOV-2002; 2002US-0423748P.

(CURA-) CURAGEN CORP.

PA Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;  
XX Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;  
PI Eilerman K, Eitenberg S, Gangolli EA, Gerlach VL, Gorman L;  
PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;  
PI Khramtsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;  
PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;  
PI Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;  
PI Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G;  
PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;  
PI Wolenc AR, Zhong M, Zhong H;  
XX

WPI; 2004-053467/05.

P-PSDB; ADH41615.

XX New NOVX polypeptides and nucleic acid molecules useful for preventing or  
DR treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
DR atherosclerosis or diabetes, in chromosome mapping, tissue typing or in  
XX pharmacogenomics.

PS Claim 20; SEQ ID NO 167; 1503pp; English.

XX The invention relates to 566 new isolated human polypeptides and their  
CC encoding genes, sequences that are at least 95% identical to these or  
CC sequences comprising one or more conservative substitutions in these. The  
CC polypeptide, polynucleotide and antibodies against the polypeptides are  
CC useful in diagnosing, treating or preventing NOVX-associated disorders,  
CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,  
CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,  
CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.  
CC The nucleic acids are further used as hybridization probes, in chromosome  
CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The  
CC polypeptides are also useful as vaccines. This sequence represents an  
CC example of the nucleic acid sequence of the invention.

SQ Sequence 6900 BP; 2323 A; 1365 C; 1297 G; 1915 T; 0 U; 0 Other;

Query Match 95.7%; Score 6799; DB 12; Length 6900;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 6858; Conservative 0; Mismatches 15; Indels 30; Gaps 2;

QY 106 ATGGATTTTCTTATCATTTTCTTTTACCTTTTATTTGGGACTTCAGAGACACAGGTGAT 165  
DB 1 ATGGATTTTCTTATCATTTTCTTTTACCTTTTATTTGGGACTTCAGAGACACAGGTGAT 60

QY 166 GTTTCCAATGTCGTTCTGTTACTAGGTACGATATACCATCTCTTCAATTTCTACAACA 225  
DB 61 GTTTCCAATGTCGTTCTGTTACTAGGTACGATATACCATCTCTTCAATTTCTACAACA 120

QY 226 TACACCTCACCTGTTACTAGATAGTGACACCAATGTAAACAAACACAGGCTCCAGTC 285  
DB 121 TACACCTCACCTGTTACTAGATAGTGACACCAATGTAAACAAACACAGGCTCCAGTC 180

QY 286 TTCCTAGCCGGGAAAGAGTCGATCTCTGGGATTTCTTCTGTTGGGAATACACACCT 345  
DB 181 TTCCTAGCCGGGAAAGAGTCGATCTCTGGGATTTCTTCTGTTGGGAATACACACCT 240

QY 346 AATCCAAATGGAAGGATTTATATCTTACATTTGCAATATAGGAAGTTTGTCCGTGATG 405  
DB 241 AATCCAAATGGAAGGATTTATATCTTACATTTGCAATATAGGAAGTTTGTCCGTGATG 300

QY 406 CAAACAGTATATACCAAGTCAGATCAAAAGCCAGACAGTCCTGGAAGTTCTTCTTACTAAT 465  
DB 301 CAAACAGTATATACCAAGTCAGATCAAAAGCCAGACAGTCCTGGAAGTTCTTCTTACTAAT 360

QY 466 CTTAATCTCTGGAACCAATATGAAATTAAGTTGCTGCTGAAACAGTCTGGCATTGGA 525  
DB 361 CTTAATCTCTGGAACCAATATGAAATTAAGTTGCTGCTGAAACAGTCTGGCATTGGA 420

QY 526 GTGTTTAGTGATCCATTTCTTCTTCAAACTGCAGAAAGTCTCCAGGAAAAGTGTGAAT 585  
DB 421 GTGTTTAGTGATCCATTTCTTCTTCAAACTGCAGAAAGTCTCCAGGAAAAGTGTGAAT 480

|    |      |   |      |      |  |      |
|----|------|---|------|------|--|------|
| QY | 586  | CTCACAGTTGAGGCTTACAAACGCTTCAGCAGTTAAAGCTGATTTGGTATTTTAACTCGGCAA   | 645  | 1666 | ATTACTGACATTCAGCTGAGTGAACAGAGCTGTCTTATGTATTATCAGGAGACTTGTACTTTCACT | 1725 |
| DB | 481  | CTCACAGTTGAGGCTTACAAACGCTTCAGCAGTTAAAGCTGATTTGGTATTTTAACTCGGCAA   | 540  | 1561 | ATTACTGACATTCAGCTGAGTGAACAGAGCTGTCTTATGTATTATCAGGAGACTTGTACTTTCACT | 1620 |
| QY | 646  | CCAAATGGCAAAATTAACCAAGCTTCAAGATTTAGTGTCAAAACATGCGCAGAAAGTGGATAGTA | 705  | 1726 | GAGCACAATGATTTAGTGTATCTGCTTTTACCATCATCTGGGAGAGGACCAACACAGTTCTC     | 1785 |
| DB | 541  | CCAAATGGCAAAATTAACCAAGCTTCAAGATTTAGTGTCAAAACATGCGCAGAAAGTGGATAGTA | 600  | 1621 | GAGCACAATGATTTAGTGTATCTGCTTTTACCATCATCTGGGAGAGGACCAACACAGTTCTC     | 1680 |
| QY | 706  | GTGAAAGATGTCTCAATACAGAGTAGAGGACATTTTTCATCTGGGAAATTTGCCAGAAATCAAT  | 765  | 1786 | AGTGTGAGCAGCTGAGCAAGTGCAGAGCTCCATTTAAATTTATTAATTAATAATATT          | 1845 |
| DB | 601  | GTGAAAGATGTCTCAATACAGAGTAGAGGACATTTTTCATCTGGGAAATTTGCCAGAAATCAAT  | 660  | 1681 | AGTGTGAGCAGCTGAGCAAGTGCAGAGCTCCATTTAAATTTATTAATTAATAATATT          | 1740 |
| QY | 766  | GAGAAATGATTTCTTTTATGAGTAGAGGACATTTTTCATCTGGGAAATTTGCCAGAAATCAAT   | 825  | 1846 | AGTGTGAGCAGCTGAGCAAGTGCAGAGCTCCATTTAAATTTATTAATTAATAATATT          | 1905 |
| DB | 661  | GAGAAATGATTTCTTTTATGAGTAGAGGACATTTTTCATCTGGGAAATTTGCCAGAAATCAAT   | 720  | 1741 | AGTGTGAGCAGCTGAGCAAGTGCAGAGCTCCATTTAAATTTATTAATTAATAATATT          | 1800 |
| QY | 826  | ACACCTTCCATCGGTTACCAACATTCATCAAGCAGCTTGACACAGAAATGAGATCAGCTCT     | 885  | 1906 | CACTATACGATTTATGCAATGGAAATGGATACAAACAGAGCAATCCAGATAACTACCATTA      | 1965 |
| DB | 721  | ACACCTTCCATCGGTTACCAACATTCATCAAGCAGCTTGACACAGAAATGAGATCAGCTCT     | 780  | 1801 | CACTATACGATTTATGCAATGGAAATGGATACAAACAGAGCAATCCAGATAACTACCATTA      | 1860 |
| QY | 886  | GTGTGAAAGAGCCTATCAGTTTGTAGTGACACATTTGAGACCTTTATACACATATCTT        | 945  | 1966 | GATAACAGCTTTTCTCATACAGGTTTAAAGAAATACACAAATACAAATGAGAGTGGCA         | 2025 |
| DB | 781  | GTGTGAAAGAGCCTATCAGTTTGTAGTGACACATTTGAGACCTTTATACACATATCTT        | 840  | 1861 | GATAACAGCTTTTCTCATACAGGTTTAAAGAAATACACAAATACAAATGAGAGTGGCA         | 1920 |
| QY | 946  | TTTGAAGTTTCAGCTGCTCACTCAAGCAGGTTTATTTGATAGTACGATTTGTCAAGACA       | 1005 | 2026 | GCCTCAACCCACGATGGAGAAAGTTCTTTGTCTGAGAAATTAACATCTTTGTGAGAACT        | 2085 |
| DB | 841  | TTTGAAGTTTCAGCTGCTCACTCAAGCAGGTTTATTTGATAGTACGATTTGTCAAGACA       | 900  | 1921 | GCCTCAACCCACGATGGAGAAAGTTCTTTGTCTGAGAAATTAACATCTTTGTGAGAACT        | 1980 |
| QY | 1006 | CCAGAAATCAGTGCCTGAAAGCACCACCAAACTGGGTAAACAGGCAACATCAAGGAAAG       | 1065 | 2086 | TCAGAAGATGAACCCGGAATCATCACTCAAGATGTGCAAGTAAATGATGTTTACCGCAGAT      | 2145 |
| DB | 901  | CCAGAAATCAGTGCCTGAAAGCACCACCAAACTGGGTAAACAGGCAACATCAAGGAAAG       | 960  | 1981 | TCAGAAGATGAACCCGGAATCATCACTCAAGATGTGCAAGTAAATGATGTTTACCGCAGAT      | 2040 |
| QY | 1066 | TCCTTTTCAATTTTATGGGACCCACCACTATAGTAAACAGGGAATTTTAGTTATAGAGTT      | 1125 | 2146 | GAATAAGGTTGAGTGTGATCACCACCCGAAAGCCCAATGGGATCATTTATGCTTATGAA        | 2205 |
| DB | 961  | TCCTTTTCAATTTTATGGGACCCACCACTATAGTAAACAGGGAATTTTAGTTATAGAGTT      | 1020 | 2041 | GAATAAGGTTGAGTGTGATCACCACCCGAAAGCCCAATGGGATCATTTATGCTTATGAA        | 2100 |
| QY | 1126 | GAATATATGGACCATCAGTGCATTTTGGATACAGCACAACAGACCTCAAGTTTGA           | 1185 | 2206 | GTGCTATATAAAAAATATAGATCTTTATATATGAAGAACACATCAACACAGACATAATA        | 2265 |
| DB | 1021 | GAATATATGGACCATCAGTGCATTTTGGATACAGCACAACAGACCTCAAGTTTGA           | 1080 | 2101 | GTGCTATATAAAAAATATAGATCTTTATATATGAAGAACACATCAACACAGACATAATA        | 2160 |
| QY | 1186 | TTCACTAACTCAACCATTTTAACTGATGATCTATATTTGCGGCTGAAACAGTGCA           | 1245 | 2266 | TTAAGGAATTAAGACCTCACAACCTCTATAACATTTCTGTAAAGGTTTACACAGATTT         | 2325 |
| DB | 1081 | TTCACTAACTCAACCATTTTAACTGATGATCTATATTTGCGGCTGAAACAGTGCA           | 1140 | 2161 | TTAAGGAATTAAGACCTCACAACCTCTATAACATTTCTGTAAAGGTTTACACAGATTT         | 2220 |
| QY | 1246 | GGGACTGGGCCCAGTCAAAATTTTCAGTATTCATCTCCACAGATTTCCAGGGGAGTG         | 1305 | 2326 | GGTCAATGGCAATCAGGTATCTTTTCTCTCTGTAAAGACTTCGGAGACTGTGCTGAT          | 2385 |
| DB | 1141 | GGGACTGGGCCCAGTCAAAATTTTCAGTATTCATCTCCACAGATTTCCAGGGGAGTG         | 1200 | 2221 | GGTCAATGGCAATCAGGTATCTTTTCTCTCTGTAAAGACTTCGGAGACTGTGCTGAT          | 2280 |
| QY | 1306 | TTTGATTTTCACTTCAGAGGTAGAAATCCAGCAAGTAAAGTAACTTTGGAGAAACCA         | 1365 | 2386 | AGTGCAACGAAATATCACTTACAAAATATTTTCTCTGGAGAGATTTGAGCTATCATTTTC       | 2445 |
| DB | 1201 | TTTGATTTTCACTTCAGAGGTAGAAATCCAGCAAGTAAAGTAACTTTGGAGAAACCA         | 1260 | 2281 | AGTGCAACGAAATATCACTTACAAAATATTTTCTCTGGAGAGATTTGAGCTATCATTTTC       | 2340 |
| QY | 1366 | CGACAAACCAATGGAATTTTAACTAATCCAGTCAAAAGTGTAGTTCAGAGACAGGA          | 1425 | 2446 | CTTCCCCCAAGTAGTCCCAATGGAATCATATAAAAAATATACAATTTTATCTCAAGGAAGT      | 2505 |
| DB | 1261 | CGACAAACCAATGGAATTTTAACTAATCCAGTCAAAAGTGTAGTTCAGAGACAGGA          | 1320 | 2341 | CTTCCCCCAAGTAGTCCCAATGGAATCATATAAAAAATATACAATTTTATCTCAAGGAAGT      | 2400 |
| QY | 1426 | ATAATTTTGGAAATATCTTTGCTCACTGGAAATTAATGAGTATATAAATGACCCATGGCT      | 1485 | 2506 | AATGGAATGAGGAAGACTATAATAACCTCTTTAAACCCCAACCAATTAAGTACTG            | 2565 |
| DB | 1321 | ATAATTTTGGAAATATCTTTGCTCACTGGAAATTAATGAGTATATAAATGACCCATGGCT      | 1380 | 2401 | AATGGAATGAGGAAGACTATAATAACCTCTTTAAACCCCAACCAATTAAGTACTG            | 2460 |
| QY | 1486 | CCAGAAATTTGAAATAGTAGAGCAATGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT      | 1545 | 2566 | AAGAAATATACCAATATATCAATGAGGTGTCTGTAGTACACTGAAAGGTGAAGAGTT          | 2625 |
| DB | 1381 | CCAGAAATTTGAAATAGTAGAGCAATGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT      | 1440 | 2461 | AAGAAATATACCAATATATCAATGAGGTGTCTGTAGTACACTGAAAGGTGAAGAGTT          | 2520 |
| QY | 1546 | TCGCTGACCTTCACTTCTGCTACATTTATATAAAGCAGCCATCCAGATAAAAACTTTT        | 1605 | 2626 | CGGAGTGTCCCATTAAGTATATGACGAGGAGAGTGTCTCTGATTTCTCCCCCTCAAGAC        | 2685 |
| DB | 1441 | TCGCTGACCTTCACTTCTGCTACATTTATATAAAGCAGCCATCCAGATAAAAACTTTT        | 1500 | 2521 | CGGAGTGTCCCATTAAGTATATGACGAGGAGAGTGTCTCTGATTTCTCCCCCTCAAGAC        | 2580 |
| QY | 1606 | CCTGCAAGGAATAGAGCTGAAGACCAAGCTTCAACAGTTGTATCAAGGAATCAGTAT         | 1665 | 2686 | TTCTCTGTAAAAACAGTTGTCTGCTGACGGTGAAGTGTGTCATGCAACACCCCTGGAG         | 2745 |
| DB | 1501 | CCTGCAAGGAATAGAGCTGAAGACCAAGCTTCAACAGTTGTATCAAGGAATCAGTAT         | 1560 | 2581 | TTCTCTGTAAAAACAGTTGTCTGCTGACGGTGAAGTGTGTCATGCAACACCCCTGGAG         | 2640 |
|    |      |   |      | 2746 | CCAAATGGAATTTATCTTTTATTAACACAGTTTATGTCTGGAATAGATCATTTAAAAAACT      | 2805 |



Db 2641 CCAATGGAAATATCCTTTATACAGATTTATGCTGGAAATAGATCATCAITTAANAAC 2700  
Qy  
Db 2806 ATTAATGTCTCACTGAACATCATTTGGAGTTATCAGATTTGGATTATTAATGTGTAATACAGT 2865  
Qy  
Db 2701 ATTAATGTCTCACTGAACATCATTTGGAGTTATCAGATTTGGATTATTAATGTGTAATACAGT 2760  
Qy  
Db 2866 GCTTATGTAAACAGCTAGCACAGATTTGGTGATGGGAAACACAGGAAGCAATATCATTTAGC 2925  
Qy  
Db 2761 GCTTATGTAAACAGCTAGCACAGATTTGGTGATGGGAAACACAGGAAGCAATATCATTTAGC 2820  
Qy  
Db 2926 TTTCAAACACAGAGGAGCAGCAGATCTCCCAAAGATGCTTTATTTATGCAAACTC 2985  
Qy  
Db 2821 TTTCAAACACAGAGGAGCAGCAGATCTCCCAAAGATGCTTTATTTATGCAAACTC 2880  
Qy  
Db 2986 AGTTCCTTCATCAATAATCTTTCTGGACACCTCTTCAAACCTTAATGGGATTAATACAA 3045  
Qy  
Db 2881 AGTTCCTTCATCAATAATCTTTCTGGACACCTCTTCAAACCTTAATGGGATTAATACAA 2940  
Qy  
Db 3046 TATTACTCTGTTTATACAGAAATATCTTCAAGTACTTTTATGAGAGATTTTACATCCAT 3105  
Qy  
Db 2941 TATTACTCTGTTTATACAGAAATATCTTCAAGTACTTTTATGAGAGATTTTACATCCAT 3000  
Qy  
Db 3106 GAACCTAACCAATGACTTTGCAATATGACTGTATCCAAATTAATGATAAAGTACATA 3165  
Qy  
Db 3001 GAAGTAAACCAATGACTTTGCAATATGACTGTATCCAAATTAATGATAAAGTACATA 3060  
Qy  
Db 3166 TTCAGCTACTATACATTTGGTTTAAACAGCAAGTACTTCAAGTGGAAATGGGAATAAAGC 3225  
Qy  
Db 3061 TTCAGCTACTATACATTTGGTTTAAACAGCAAGTACTTCAAGTGGAAATGGGAATAAAGC 3120  
Qy  
Db 3226 AGTGACATCAATGAAGTATACAGATCAAGATCAAGATCAAGTGGGTTGTTGGAACCTG 3285  
Qy  
Db 3121 AGTGACATCAATGAAGTATACAGATCAAGATCAAGATCAAGTGGGTTGTTGGAACCTG 3180  
Qy  
Db 3286 ACTTACGAATCCATTTGCTCACTCAATTAATGTAAGTGGTCCCAAGGCTCAACCA 3345  
Qy  
Db 3181 ACTTACGAATCCATTTGCTCACTCAATTAATGTAAGTGGTCCCAAGGCTCAACCA 3240  
Qy  
Db 3346 AACGGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGCAGACTCTCGCCATGTGAGA 3405  
Qy  
Db 3241 AACGGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGCAGACTCTCGCCATGTGAGA 3300  
Qy  
Db 3406 CCACTCTTGTATACATATGAGAGAGCATATATTTTGATAATCTGGAATAACACTGAT 3465  
Qy  
Db 3301 CCACTCTTGTATACATATGAGAGAGCATATATTTTGATAATCTGGAATAACACTGAT 3360  
Qy  
Db 3466 TATATATTAATAATTTACTCCATCAAGAAAGGATCTCTGATACCTATACCTGCCAG 3525  
Qy  
Db 3361 TATATATTAATAATTTACTCCATCAAGAAAGGATCTCTGATACCTATACCTGCCAG 3420  
Qy  
Db 3526 CTATACATCAAGACTGAAGAGATGTCAGAAACTTCCAAATATCAACACTTTTAA 3585  
Qy  
Db 3421 CTATACATCAAGACTGAAGAGATGTCAGAAACTTCCAAATATCAACACTTTTAA 3480  
Qy  
Db 3586 AACCTTCTCTACTCCTCAGTCTCTTATCATGAGATCCCAAGTAAAGGCAATGTTGCA 3645  
Qy  
Db 3481 AACCTTCTCTACTCCTCAGTCTCTTATCATGAGATCCCAAGTAAAGGCAATGTTGCA 3540  
Qy  
Db 3646 ATAAATGTTATGAATTAATTTTACAGGACCAATGAATAATTTCTTTTCAATTAATCTT 3705  
Qy  
Db 3541 ATAAATGTTATGAATTAATTTTACAGGACCAATGAATAATTTCTTTTCAATTAATCTT 3600  
Qy  
Db 3706 GATAATTAATATATTTTGAAGAGCTTCCCAATTTTACATTAATAGCTTTTGTGCTGC 3765  
Qy  
Db 3601 GATAATTAATATATTTGAAGAGCTTCCCAATTTTACATTAATAGCTTTTGTGCTGC 3660  
Qy  
Db 3766 GCAAGAACTAGAAAAGGACTTGTCTCTCCAGTATTTCTTTTACAGATGAGTCA 3825  
Qy  
Db 3661 GCAAGAACTAGAAAAGGACTTGTCTCTCCAGTATTTCTTTTACAGATGAGTCA 3720  
Qy  
Db 3826 GTCCGTTAGCACTCCAAATAATTTGACTTTTAACTCAACTGATCTTACAGCTTTGATGG 3885

Db 3721 GTCCGTTAGCACTCCAAATAATTTGACTTTAATCAACTGACTTACAGCTTTGTATGG 3780  
Qy  
Db 3886 CTGAAATGGAGCCCAAGTCTCTTCCAGTGGTATTTGTTAAAGTATATATAGTTTAAAT 3945  
Qy  
Db 3781 CTGAAATGGAGCCCAAGTCTCTTCCAGTGGTATTTGTTAAAGTATATATAGTTTAAAT 3840  
Qy  
Db 3946 CATGAACATGAACCTGACACTATATATTAAGATATATCAGGATTTAAACTGAGCC 4005  
Qy  
Db 3841 CATGAACATGAACCTGACACTATATATTAAGATATATCAGGATTTAAACTGAGCC 3900  
Qy  
Db 4006 AAACCTGTTGGACTGGAACAGTCAGCACTCTCTATCCGTGTATCTGCGTTTCAACAA 4065  
Qy  
Db 3901 AAACCTGTTGGACTGGAACAGTCAGCACTCTCTATCCGTGTATCTGCGTTTCAACAA 3960  
Qy  
Db 4066 GTTGGAAATGGCAATCAATTTAGTAAATAGTAAATTAACAAACCAAGATCAGTTCCA 4125  
Qy  
Db 3961 GTTGGAAATGGCAATCAATTTAGTAAATAGTAAATTAACAAACCAAGATCAGTTCCA 4020  
Qy  
Db 4126 GATGTCGTGAGATATGCAAGTGGCAACTAGCTGGCAGTCAGTTTGTAGTAATGG 4185  
Qy  
Db 4021 GATGTCGTGAGATATGCAAGTGGCAACTAGCTGGCAGTCAGTTTGTAGTAATGG 4080  
Qy  
Db 4186 GATCCACCCAAAGGCAAAATGGAATAATAACGAGTATATGTAACAGTTGAAAGGAAT 4245  
Qy  
Db 4081 GATCCACCCAAAGGCAAAATGGAATAATAACGAGTATATGTAACAGTTGAAAGGAAT 4140  
Qy  
Db 4246 TCTCAAAAAGTTTCTCCCAAGATCACTGTACATTTTCAATAAGCTTCTTGGCAATACC 4305  
Qy  
Db 4141 TCTCAAAAAGTTTCTCCCAAGATCACTGTACATTTTCAATAAGCTTCTTGGCAATACC 4200  
Qy  
Db 4306 TCATATGCTTTTAAAGTAAAGGCTTCAACCTCAGCTGGTGAAGGTGATGAAGACATGC 4365  
Qy  
Db 4201 TCATATGCTTTTAAAGTAAAGGCTTCAACCTCAGCTGGTGAAGGTGATGAAGACATGC 4260  
Qy  
Db 4366 CATGTCAGCACACTCCTGAAAAGTCCCAAGTGTCCCAAAATATGCTTTTCTGAT 4425  
Qy  
Db 4261 CATGTCAGCACACTCCTGAAAAGTCCCAAGTGTCCCAAAATATGCTTTTCTGAT 4320  
Qy  
Db 4426 GTTCAGTCAACTAGTGCAACATTGACATGATGAAGACCTGACACTATCTTGGCTACTTT 4485  
Qy  
Db 4321 GTTCAGTCAACTAGTGCAACATTGACATGATGAAGACCTGACACTATCTTGGCTACTTT 4380  
Qy  
Db 4486 CAAAATTAACAAATTTACCACTCACTTCTGCTGCTCAAAAATGCAAGATGGGAATCCGAA 4545  
Qy  
Db 4381 CAAAATTAACAAATTTACCACTCACTTCTGCTGCTCAAAAATGCAAGATGGGAATCCGAA 4440  
Qy  
Db 4546 GAATGTTGTAATATCAAAAATTTCAATACCTCTATGAAGCTCACCTTAACTGAAGAGACA 4605  
Qy  
Db 4441 GAATGTTGTAATATCAAAAATTTCAATACCTCTATGAAGCTCACCTTAACTGAAGAGACA 4500  
Qy  
Db 4606 GTATATGGAATTAAGAAATTTTAGATGGTATAGATTTCAAGTGGCTGCCAGCAATGCT 4665  
Qy  
Db 4501 GTATATGGAATTAAGAAATTTTAGATGGTATAGATTTCAAGTGGCTGCCAGCAATGCT 4560  
Qy  
Db 4666 GGCTATGGCAATGCTTTCAAACTGGATTTCTCAAAAATCTCTGCTGGCCCTCCAGATGGT 4725  
Qy  
Db 4561 GGCTATGGCAATGCTTTCAAACTGGATTTCTCAAAAATCTCTGCTGGCCCTCCAGATGGT 4620  
Qy  
Db 4726 CCTCTGGAATTTCTGATGATGAGCAATCACTTTTGTAGCATCAGATTAAGCTGGAGT 4785  
Qy  
Db 4621 CCTCTGGAATTTCTGATGATGAGCAATCACTTTTGTAGCATCAGATTAAGCTGGAGT 4680  
Qy  
Db 4786 GAACCTGCTGCTCAATTTACTGGAACCAACATGTTATCTGATTTAGTGTCAAAATCGGTAGATAAT 4845  
Qy  
Db 4681 GAACCTGCTGCTCAATTTACTGGAACCAACATGTTATCTGATTTGATGTCAAAATCGGTAGATAAT 4740  
Qy  
Db 4846 GATGAATTTAATATATCTTCTCATCAAGTCAAAATGAAGAAATAAACCATAGAAATTA 4905  
Qy  
Db 4741 GATGAATTTAATATATCTTCTCATCAAGTCAAAATGAAGAAATAAACCATAGAAATTA 4800  
Qy  
Db 4906 GATTTAGAAATTAATCAAGGATTTCTGTAGTATGATCACTGCAATTTACTGGAACATAGT 4965  
Qy  
Db 4801 GATTTAGAAATTAATCAAGGATTTCTGTAGTATGATCACTGCAATTTACTGGAACATAGT 4860





|    |      |   |      |  |      |   |  |      |  |
|----|------|---|------|--|------|---|--|------|--|
| QY | 1360 | AAACACGCAACCAAAATGGAATTAATAACCAATACCGAGTGAAAGTCTAGTTCACGAG      | 1419 |  | 2338 | GAGCTATCATTCCTTCCGCCAAGTAGTCCCAATGGAATCATACAAAAATATACAAATTTAT | 2397   |      |  |
| Db | 1261 | AAACACGCAACCAAAATGGAATTAATAACCAATACCGAGTGAAAGTCTAGTTCACGAG      | 1320 |  | QY   | 2494  | CTCAGAGAAAGTAATGGAATGCGAAGAAAGAACTATAAATACAACTCTTTTAAACCCAAAAC     | 2553 |  |
| QY | 1420 | ACAGGAATTAATTTGGAAATATCTTTGCTCACTGGAATAAATGAGATATATAATGACCCC    | 1479 |  | Db   | 2398  | CTCAAGAGAAAGTAATGGAATGCGAAGAAAGAACTATAAATACAACTCTTTTAAACCCAAAAC    | 2457 |  |
| Db | 1321 | ACAGGAATTAATTTGGAAATATCTTTGCTCACTGGAATAAATGAG---ATAAATGACCCC    | 1377 |  | QY   | 2554  | ATTTAAAGTACTGAAGAAATATACCCAAATATATCATTTGAGGTCTCTGCTAGTAGTACACTGAAA | 2613 |  |
| QY | 1480 | ATGGCTCCAGAAATCTGGAACATAGTAGAGCCCAATGGTAGGATATATGAGGGTTCAGCA    | 1539 |  | Db   | 2458  | ATT-----CTGAGAAATATACCCAAATATATCATTTGAGGTCTCTGCTAGTAGTACACTCAA     | 2511 |  |
| Db | 1378 | ATGGCTCCAGAAATCTGGAACATAGTAGAGCCCAATGGTAGGATATATGAGGGTTCAGCA    | 1437 |  | QY   | 2614  | GGTGAAGGAGTTTGGAGTCTCCCAATAAGTATATCTGAAGGAAAGATGCTCCTGATTTCT       | 2673 |  |
| QY | 1540 | GAGATGCTGCTGACCTTCACTCACTTGCTACATTTATATATAACAGCCCATCCAGATAAA    | 1599 |  | Db   | 2512  | GGTGAAGGAGTTTGGAGTCTCCCAATAAGTATATCTGAAGGAAAGATGCTCCTGATTTCT       | 2571 |  |
| Db | 1438 | GAGATGCTGCTGACCTTCACTCACTTGCTACATTTATATATAACAGCCCATCCAGATAAA    | 1497 |  | QY   | 2674  | CCCCCTCAAGACTTCTGTATAAAACAGTTGCTGTGTGTCACGGTCAAGTTGTTCATGGCAA      | 2733 |  |
| QY | 1600 | AACCTTTCTGCAAGGAATAGAGTGAAGACCAAGCTTCCACGATTTGTAATCAAGGAAT      | 1659 |  | Db   | 2572  | CCCCCTCAAGACTTCTGTATAAAACAGTTGCTGTGTGTCACGGTCAAGTTGTTCATGGCAA      | 2631 |  |
| Db | 1498 | AACCTTTCTGCAAGGAATAGAGTGAAGACCAAGCTTCCACGATTTGTAATCAAGGAAT      | 1557 |  | QY   | 2734  | CCACCCCTGGAGCCAAATGGAATTAATCTTTTATTAACACAGTTTATGTCT---GGAATAGA     | 2790 |  |
| QY | 1660 | CAGTATATTAATGACATTCGAGCTGAACAGCTGCTTATGTTATCAGGAGACTTGTACCT     | 1719 |  | Db   | 2632  | CCACCCCTGGAGCCAAATGGAATTAATCTTTTATTAACACAGTTTATGTCTGGAAGGAATAGA    | 2691 |  |
| Db | 1558 | CAGTATATTAATGACATTCGAGCTGAACAGCTGCTTATGTTATCAGGAGACTTGTACCT     | 1617 |  | QY   | 2791  | TCATCATTTAAAAAATTAATATCTCACTGAAACATCATTTGGAGTTATCAGATTTGGATTTAT    | 2850 |  |
| QY | 1720 | TTCACTGAGCACATGATTAAGTGTATCTGCTTTTCAACCATCATGGGAGAGGACCCACCAACA | 1779 |  | Db   | 2692  | TCATCATTTAAAAAATTAATATCTCACTGAAACATCATTTGGAGTTATCAGATTTGGATTTAT    | 2751 |  |
| Db | 1618 | TTCACTGAGCACATGATTAAGTGTATCTGCTTTTCAACCATCATGGGAGAGGACCCACCAACA | 1677 |  | QY   | 2851  | AAATGTTGAATACAGTCTTATGTAACGCTAGCACACAGATTTGGTGTGAGGAAACACAGA       | 2910 |  |
| QY | 1780 | GTTCTCAGTGTTAGGACAGCTCAGCAAGTGCCAAAGCTCCATTAAAAATTAATACTATAAA   | 1839 |  | Db   | 2752  | AAATGTTGAATACAGTCTTATGTAACAGCTAGCACACAGATTTGGTGTGAGGAAACACAGA      | 2811 |  |
| Db | 1678 | GTTCTCAGTGTTAGGACAGCTCAGCAAGTGCCAAAGCTCCCAAGATTTCCCAATGGAAAA    | 1737 |  | QY   | 2911  | AGCAATATCATTTAGCTTTTCAAAACACAGAGGAGAGCACCAAGCGATCTCCCAAGATGTT      | 2970 |  |
| QY | 1840 | AATATTAGTCTTCATCTATTGTTGTTATATGGAGTCTCCAGAAATATCCCAATGAAAA      | 1899 |  | Db   | 2812  | AGCAATATCATTTAGCTTTTCAAAACACAGAGG---ACCAAGCGATCTCCCAAGATGTT        | 2868 |  |
| Db | 1738 | AATATTAGTCTTCATCTATTGTTGTTATATGGAGTCTCCAGAAATATCCCAATGGAAAA     | 1797 |  | QY   | 2971  | TATTATGCAAAACCTCAGTTCTTTCATCAATAAATTTCTTTCTGACACCTCTCTTCAAAACCT    | 3030 |  |
| QY | 1900 | ATAACTCAGTATACGATTTATGCAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT   | 1959 |  | Db   | 2869  | TATTATGCAAAACCTCAGTTCTTTCATCAATAAATTTCTTTCTGACACCTCTCTTCAAAACCT    | 2928 |  |
| Db | 1798 | ATAACTCAGTATACGATTTATGCAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT   | 1857 |  | QY   | 3031  | AAATGGAATTAACAATATATCTGTTTATTAACAGAAATCTTCAGGTACTTTTATGTCAG        | 3090 |  |
| QY | 1960 | ACCATAGATAACAGCTTTCTCATPAAC-----AGGGTTAAAGAAATATACAAATAACAA     | 2013 |  | Db   | 2929  | AAATGGAATTAACAATATATCTGTTTATTAACAGAAATCTTCAGGTACTTTTATGTCAG        | 2988 |  |
| Db | 1858 | ACCATAGATAACAGCTTTCTCATPAAC-----AGGGTTAAAGAAATATACAAATAACAA     | 1917 |  | QY   | 3091  | AAATTTTACACTCCATGAACTAACCAATGACTTTTGAACAATATGACTGTATCCCAATATA      | 3150 |  |
| QY | 2014 | ATGAGAGTGGCAGCTTCAACCCAGATGGAAGATTTCTTGTCTGAAGAAATGACATC        | 2073 |  | Db   | 2989  | AAATTTTACACTCCATGAAAGTAAACCAATGACTTTTGAACAATATGACTGTATCCCAATATA    | 3048 |  |
| Db | 1918 | ATGAGAGTGGCAGCTTCAACCCAGATGGAAGATTTCTTGTCTGAAGAAATGACATC        | 1977 |  | QY   | 3151  | GATAAATGCAATAATTCAGCTATCATATACATTTTGGTTAAACAGCAAGTACTTCAGTTGGA     | 3210 |  |
| QY | 2074 | TTTGTGAGAACTTCAGAAATGAAACCGGAATCATCACCTCAAGATGTCGAAGTAAATGAT    | 2133 |  | Db   | 3049  | GATAAATGCAATAATTCAGCTATCATATACATTTTGGTTAAACAGCAAGTACTTCAGTTGGA     | 3108 |  |
| Db | 1978 | TTTGTGAGAACTTCAGAAATGAAACCGGAATCATCACCTCAAGATGTCGAAGTAAATGAT    | 2037 |  | QY   | 3211  | AAATGGAATTAAGCAAGTGAACATCATTTGAAGTATACACAGATCAAGACATCTGGAAGGG      | 3270 |  |
| QY | 2134 | GTTTACCGCAGATGAAATAAGGTTGAAGTGTCACCAACCGGAAAGCCCAATGGATCATTT    | 2193 |  | Db   | 3109  | AAATGGAATTAAGCAAGTGAACATCATTTGAAGTATACACAGATCAAGACCTCCCTGAAGGG     | 3168 |  |
| Db | 2038 | GTTTACCGCAGATGAAATAAGGTTGAAGTGTCACCAACCGGAAAGCCCAATGGATCATTT    | 2097 |  | QY   | 3271  | TTTGTGGAACCTGACTTTACGAATCCATTTTCGTCAAATGCAATAAATGTAAGCTGGGTC       | 3330 |  |
| QY | 2194 | ATTGCTTTAGAGTGTCTATATAAATAATATAGATACATTTTATATATGAAGACATCAACA    | 2253 |  | Db   | 3169  | TTTGTGGAACCTGACTTTACGAATCCATTTTCGTCAAATGCAATAAATGTAAGCTGGGTC       | 3228 |  |
| Db | 2098 | ATTGCTTTAGAGTGTCTATATAAATAATATAGATACATTTTATATGAAGACATCAACA      | 2157 |  | QY   | 3331  | CCAACCGGCTCAAACCAACGGTCTAGTCTTCTACTATGTTTCTACTGATCTTACAGCAGACT     | 3390 |  |
| QY | 2254 | ACAGACATATAATTAAGGAATTAAGACCTTCAACCTCTATATAACATTTCTGTAAGGTCT    | 2313 |  | Db   | 3229  | CCAACCGGCTCAAACCAACGGTCTAGTCTTCTACTATGTTTCTACTGATCTTACAGCAGACT     | 3288 |  |
| Db | 2158 | ACAGACATATAATTAAGGAATTAAGACCTTCAACCTCTATATAACATTTCTGTAAGGTCT    | 2217 |  | QY   | 3391  | CTTCGCCATGTGAGACCACTCTTGTATACATATGAGAGGACATATTTTGTATATCTG          | 3450 |  |
| QY | 2314 | TACACAGATTTGGTCAATGCAATCAGGTATCTTTTACTCTCTGTAAGGACTTCGGAG       | 2373 |  | Db   | 3289  | CTTCGCCATGTGAGACCACTCTTGTATACATATGAGAGGACATATATTTTGTATAATCTG       | 3348 |  |
| Db | 2218 | TACACAGATTTGGTCAATGCAATCAGGTATCTTTTACTCTCTGTAAGGACTTCGGAG       | 2277 |  | QY   | 3451  | GAATAATACACTGATTATATATATAAATAATTAATCTCAATCAACAGAAAGGATTTCTCTGAT    | 3510 |  |
| QY | 2374 | ACTGTGCTGATAGTGCACAGAAATATCACTTACAAAAATTTCTTCTGAGAGATTT         | 2433 |  | Db   | 3349  | GAATAATACACTGATTATATATAAATAATTAATCTCAATCAACAGAAAGGATTTCTCTGAT      | 3408 |  |
| Db | 2278 | ACTGTGCTGATAGTGCACAGAAATATCACTTACAAAAATTTCTTCTGAGAGATTT         | 2337 |  | QY   | 3511  | ACCTATCTGCCAGCTATATCATCAAGACTGAAGAGATGTCCAGAAATCTTCCACCAATA        | 3570 |  |
| QY | 2434 | GAGCTATCATTCCTTCCCCCAAGTAGTCCCAATGGAATCATAAAAAATATACAAATTTAT    | 2493 |  |      |   |  |      |  |









|    |      |  |      |    |      |   |      |
|----|------|--|------|----|------|---|------|
| QY | 1240 | AGTCAGGAGCTGGGCCCAAGTCAAAATATTTTCAGTATTCATCTCCACAGATGTTCCAGGG  | 1299 | Db | 2218 | TACACAGATTTGGTCAATGCGCAATCAGGTATCTCTTTACTCTCTGTAGGAGCTTCGGAG      | 2277 |
| Db | 1141 | AGTCAGGAGCTGGGCCCAAGTCAAAATATTTTCAGTATTCATCTCCACAGATGTTCCAGGG  | 1200 | QY | 2374 | ACTGTGCTGTAGTGCACCGAATAATATCCTTACCAAAATATTTCTCTCGAGAGATT          | 2433 |
| QY | 1300 | GCAGTGTGTTGATTTACAACTTGCGAGGTAGAAATCACGCAAGTAAAGATTTCTTGGAAG   | 1359 | Db | 2278 | ACTGTGCTGTAGTGCACCGAATAATATCCTTACCAAAATATTTCTCTCGAGAGATT          | 2337 |
| Db | 1201 | GCAGTGTGTTGATTTACAACTTGCGAGGTAGAAATCACGCAAGTAAAGATTTCTTGGAAG   | 1260 | QY | 2434 | GAGCTATCAATCTCTTCCCAAGTGTCCCAATGAAATCATATAAAATAATATACAAATTTAT     | 2493 |
| QY | 1360 | AAACACGCAACCAATGGAATTAATACCAATACCGAGTGAAGTGTCTAGTTCAGAG        | 1419 | Db | 2338 | GAGCTATCAATCTCTTCCCAAGTGTCCCAATGAAATCATATAAAATAATATACAAATTTAT     | 2397 |
| Db | 1261 | AAACACGCAACCAATGGAATTAATACCAATACCGAGTGAAGTGTCTAGTTCAGAG        | 1320 | QY | 2494 | CTCAAGAGAGTAAATGGAATCAGGAAAGAACTATAAAATACAACTCTTTTAAACCCAAAAC     | 2553 |
| QY | 1420 | ACAGGAATATTTTGGAATACTCTTGCTCACTGGAATAATAGTATATAATGACCC         | 1479 | Db | 2398 | CTCAAGAGAGTAAATGGAATCAGGAAAGAACTATAAAATACAACTCTTTTAAACCCAAAAC     | 2457 |
| Db | 1321 | ACAGGAATATTTTGGAATACTCTTGCTCACTGGAATAATAGTATATAATGACCC         | 1377 | QY | 2554 | ATTAAAGTACTGAAGAAATATACCAATATATCATTTGAGGTCTGTCTAGTACACTGAA        | 2613 |
| QY | 1480 | ATGGCTCCAGAAATTTGGAACATAGTAGAGCCAAATGGTAGGATTTATATGAGGTTTCAGCA | 1539 | Db | 2458 | ATT-----CTGAAGAAATATATACCAATATATCATTTGAGGTCTGTCTAGTACACTGAA       | 2511 |
| Db | 1378 | ATGGCTCCAGAAATTTGGAACATAGTAGAGCCAAATGGTAGGATTTATATGAGGTTTCAGCA | 1437 | QY | 2614 | GGTGAAGGAGTTTCGAGTGTCTCCCATAGTATACCTGACGAGGAAAGATGCTCTGATTTCT     | 2673 |
| QY | 1540 | GAGATGTGCTGACCTTCACTCACTGTCTACATTTATATATAACGCCATCCAGATAA       | 1599 | Db | 2512 | GGTGAAGGAGTTTCGAGTGTCTCCCATAGTATACCTGACGAGGAAAGATGCTCTGATTTCT     | 2571 |
| Db | 1438 | GAGATGTGCTGACCTTCACTCACTGTCTACATTTATATAACGCCATCCAGATAA         | 1497 | QY | 2674 | CCCCCTCAAGACTTCTCTGTAAAAACAGTTGTCTGGTGTCAACGGTGAAGTTGTCAATGCGCAA  | 2733 |
| QY | 1600 | AACCTTTCTGCAAGGAATAGAGCTGAAGACCAAGCTTCCACAGTTCTAACTACAAGGAAT   | 1659 | Db | 2572 | CCCCCTCAAGACTTCTCTGTAAAAACAGTTGTCTGGTGTCAACGGTGAAGTTGTCAATGCGCAA  | 2631 |
| Db | 1498 | AACCTTTCTGCAAGGAATAGAGCTGAAGACCAAGCTTCCACAGTTCTAACTACAAGGAAT   | 1557 | QY | 2734 | CAACCCCTGGAGCCAAATGGAATTTCTTTTATATACAGTTTATGTCT--GGATAGA          | 2790 |
| QY | 1660 | CAGTATATTACTGACATTTGAGCTGAACAGCTGTCTTATGTTATCAGGAGCTTGTACCT    | 1719 | Db | 2632 | CAACCCCTGGAGCCAAATGGAATTTCTTTTATATACAGTTTATGTCTGGAGGAAATAGA       | 2691 |
| Db | 1558 | CAGTATATTACTGACATTTGAGCTGAACAGCTGTCTTATGTTATCAGGAGCTTGTACCT    | 1617 | QY | 2791 | TCATCATTAATAAACTATTAAATGTCTCTGAAACATCATTTGGAGTTATCAGATTTGGATTAT   | 2850 |
| QY | 1720 | TTCACTGAGCAGATGATGATGTCTCTTTTCACTCACTGAGGAGGAGGACCAACA         | 1779 | Db | 2692 | TCATCATTAATAAACTATTAAATGTCTCTGAAACATCATTTGGAGTTATCAGATTTGGATTAT   | 2751 |
| Db | 1618 | TTCACTGAGCAGATGATGATGTCTCTTTTCACTCACTGAGGAGGAGGACCAACA         | 1677 | QY | 2851 | AATGTTGAATAACAGTGTCTTATTAACAGCTAGCAACAGATTTGGTGTATGGGAAACAGAA     | 2910 |
| QY | 1780 | GTCTCAGTGTAGGACAGCTGAGCAAGTGCACAGCTTCCATTAATAATTAATACTATAA     | 1839 | Db | 2752 | AATGTTGAATAACAGTGTCTTATTAACAGCTAGCAACAGATTTGGTGTATGGGAAACAGAA     | 2811 |
| Db | 1678 | GTCTCAGTGTAGGACAGCTGAGCAAGTGCACAGCTTCCATTAATAATTAATACTATAA     | 1737 | QY | 2911 | AGCAATATCATTTAGCTTTTCAACACAGGAGGAGGACCAAGCGATCTCCCAAGATGTT        | 2970 |
| QY | 1840 | AATPATTAGTCTTCACTAATTTTGTATATGTTGGATCTTCCAGAAATATCCCAATGGA     | 1899 | Db | 2812 | AGCAATATCATTTAGCTTTTCAACACAGGAGGACCAAGCGATCTCCCAAGATGTT           | 2868 |
| Db | 1738 | AATPATTAGTCTTCACTAATTTTGTATATGTTGGATCTTCCAGAAATATCCCAATGGA     | 1797 | QY | 2971 | TATTTATGAAACCTCAGTGTCTTCAATCAATATTTCTTTCTGACACCTCTTCAAAACCT       | 3030 |
| QY | 1900 | ATAACTCATATACAGATTTATGCAATGGAATGGATGATACAAACAGAGCATTCAGATACT   | 1959 | Db | 2869 | TATTTATGAAACCTCAGTGTCTTCAATCAATATTTCTTTCTGACACCTCTTCAAAACCT       | 2928 |
| Db | 1798 | ATAACTCATATACAGATTTATGCAATGGAATGGATGATACAAACAGAGCATTCAGATACT   | 1857 | QY | 3031 | AATGGGATTTATACAAATTTACTCTGTTTATTAACAGAAATACCTTCAAGGTACTTTTATGCGAG | 3090 |
| QY | 1960 | ACCATAGATAACAGCTTTCTCATAAC-----AGGGTTAAAGAAATACACAAATACAAA     | 2013 | Db | 2929 | AATGGGATTTATACAAATTTACTCTGTTTATTAACAGAAATACCTTCAAGGTACTTTTATGCGAG | 2988 |
| Db | 1858 | ACCATAGATAACAGCTTTCTCATAAC-----AGGGTTAAAGAAATACACAAATACAAA     | 1917 | QY | 3091 | AATTTTACACTCCATGAACTAACCAATGACTTTTGACAAATATGACTGTATCCCAATATA      | 3150 |
| QY | 2014 | ATGAGAGTGGCAGCTCAACCCCGATGGAGAAAGTTCTTTGTCTGAAGAAATGACATC      | 2073 | Db | 2989 | AATTTTACACTCCATGAACTAACCAATGACTTTTGACAAATATGACTGTATCCCAATATA      | 3048 |
| Db | 1918 | ATGAGAGTGGCAGCTCAACCCCGATGGAGAAAGTTCTTTGTCTGAAGAAATGACATC      | 1977 | QY | 3151 | GATAAACTACAAATATTCAGTACTATACATTTTGGTTAAACAGCAAGTACTTCAGTTGGA      | 3210 |
| QY | 2074 | TTTGTGAGAACTTCAGAGATGAAACCGGAATCATACCTCAAGATGTCGAAGTAAATGAT    | 2133 | Db | 3049 | GATAAACTACAAATATTCAGTACTATACATTTTGGTTAAACAGCAAGTACTTCAGTTGGA      | 3108 |
| Db | 1978 | TTTGTGAGAACTTCAGAGATGAAACCGGAATCATACCTCAAGATGTCGAAGTAAATGAT    | 2037 | QY | 3211 | AATGGGAATAAAGCAGTGCACATCATTTGAAAGTATACACAGATCAAGACATACCTGAAGGG    | 3270 |
| QY | 2134 | GTATACCGCAGATGAATTAAGGTTGAAGTGTGACACCCGAAAGCCCAATGGGATCAT      | 2193 | Db | 3109 | AATGGGAATAAAGCAGTGCACATCATTTGAAAGTATACACAGATCAAGACGCTCCCTGAAGGG   | 3168 |
| Db | 2038 | GTATACCGCAGATGAATTAAGGTTGAAGTGTGACACCCGAAAGCCCAATGGGATCAT      | 2097 | QY | 3271 | TTTGTGGAAACCTGACTTACGAAATCCATTTGGTCACTGCAATGCAATGCTTAAGCTCGGTC    | 3330 |
| QY | 2194 | ATTGCTTTATGAAGTGTATATAAATAATAGATATCTTTTATATATGAAGAAACATCAACA   | 2253 | Db | 3169 | TTTGTGGAAACCTGACTTACGAAATCCATTTGGTCACTGCAATGCAATGCTTAAGCTCGGTC    | 3228 |
| Db | 2098 | ATTGCTTTATGAAGTGTATATAAATAATAGATATCTTTTATATATGAAGAAACATCAACA   | 2157 | QY | 3331 | CCACCGGCTCAACCAAAACGGTCTAGTCTTCTACTATGTTTCACTGATCTTTACAGCAGACT    | 3390 |
| QY | 2254 | ACAGACATATATTAAGGAATTAAGACCTTCAACCCCTCTATATACTTTCTGTAAGGTCT    | 2313 | Db | 3229 | CCACCGGCTCAACCAAAACGGTCTAGTCTTCTACTATGTTTCACTGATCTTTACAGCAGACT    | 3288 |
| Db | 2158 | ACAGACATATATTAAGGAATTAAGACCTTCAACCCCTCTATATACTTTCTGTAAGGTCT    | 2217 | QY | 3391 | CCTCGCCATGTGAGACCACTCTTGTGTTAATATGAGAGAGGAAATATATTTTGTATATCTG     | 3450 |
| QY | 2314 | TACACAGATTTGGTCAATGCGCAATCAGGTATCTTCTTTACTCTCTGTAAGGAGCTTCGGAG | 2373 |    |      |   |      |

Db 3289 CCTCGCATGTGAGACCACTCTTGTGTACATATGAGAGACATATATTTTGATATCTG 3348  
Qy 3451 GAAAAATPACACTGATATATATTTAAATAATTTACATCAACAGAAAAGGATTTCTGTGAT 3510  
Db 3349 GAAAAATPACACTGATATATATTTAAATAATTTACATCAACAGAAAAGGATTTCTGTGAT 3408  
Qy 3511 ACCTATACCTGCCAGCTATACATCAAGCTGAGAGAGATGTCACAGAACTTCCACCAATA 3570  
Db 3409 ACCTATACCTGCCAGCTATACATCAAGCTGAGAGAGATGTCACAGAACTTCCACCAATA 3468  
Qy 3571 ATCAACACATTTTAAAAACCTTCTCTACCTCAGTTCTCTTATCATGGGATCCCCAGTA 3630  
Db 3469 ATCAACACATTTTAAAAACCTTCTCTACCTCAGTTCTCTTATCATGGGATCCCCAGTA 3528  
Qy 3631 AAGCCAAATGGTGCATATAAAGTTATGATTTTAACTTTTAAAGGACCAATGAAATATAT 3690  
Db 3529 AAGCCAAATGGTGCATATAAAGTTATGATTTTAACTTTTAAAGGACCAATGAAATATAT 3588  
Qy 3691 TCTTTTCATTTCTGTATTAATACATATATTTGGAAGAGCTTTTCCACATTTACATTTAT 3750  
Db 3589 TCTTTTCATTTCTGTATTAATACATATATTTGGAAGAGCTTTTCCACATTTACATTTAT 3648  
Qy 3751 AGCTTTTCTGCTCCGCAAGAACTAGAAAAGGACTTGGTCTTCCAGTATTTCTTTCTTT 3810  
Db 3649 AGCTTTTCTGCTCCGCAAGAACTAGAAAAGGACTTGGTCTTCCAGTATTTCTTTCTTT 3708  
Qy 3811 TACACAGATGAGTCAGTGGCGTTAGCACCTTCCACAAAATTTGACTTTTAACTCACTGACT 3870  
Db 3709 TACACAGATGAGTCAGTGGCGTTAGCACCTTCCACAAAATTTGACTTTTAACTCACTGACT 3768  
Qy 3871 TCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGGTATTTGTTAAAGTA 3930  
Db 3769 TCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGGTATTTGTTAAAGTA 3828  
Qy 3931 TATAGTTTAAATTCATGAACATGAACACTGACACTATATATTAAGATATATCAGGA 3990  
Db 3829 TATAGTTTAAATTCATGAACATGAACACTGACACTATATATTAAGATATATCAGGA 3888  
Qy 3991 TTTTAAACTGAAAGCCAACTTGTGGACTGGAAACCAAGTCAGCACTTACTCTATCCGTGTA 4050  
Db 3889 TTTTAAACTGAAAGCCAACTTGTGGACTGGAAACCAAGTCAGCACTTACTCTATCCGTGTA 3948  
Qy 4051 TCTGGCTTCCAAAGTTGGAATGGCAATCAATTTAGTAAATGATAGTAAATTTCACAACC 4110  
Db 3949 TCTGGCTTCCAAAGTTGGAATGGCAATCAATTTAGTAAATGATAGTAAATTTCACAACC 4008  
Qy 4111 CAAGAAATCAGTTCAGATGTCGTGAGAAATATGCAAGTGCATGCACTAGCTGGCAGTCA 4170  
Db 4009 CAAGAAATCAGTTCAGATGTCGTGAGAAATATGCAAGTGCATGCACTAGCTGGCAGTCA 4068  
Qy 4171 GTTTTGTGAAATGGGATCCACCACAAAAGGCAAAATGGAAATATAACGCAATATATGGTA 4230  
Db 4069 GTTTTGTGAAATGGGATCCACCACAAAAGGCAAAATGGAAATATAACGCAATATATGGTA 4128  
Qy 4231 ACAGTTGAAAGGAAATTTCTACAAAGTTTCTCCCAAGATCAGATGACATTTTCAATAAG 4290  
Db 4129 ACAGTTGAAAGGAAATTTCTACAAAGTTTCTCCCAAGATCAGATGACATTTTCAATAAG 4188  
Qy 4291 CTTCTTGCCAAATACCTCATATGTCCTTAAAGTAAAGCTTCAACCTCAGCTGGTGAAGGT 4350  
Db 4189 CTTCTTGCCAAATACCTCATATGTCCTTAAAGTAAAGCTTCAACCTCAGCTGGTGAAGGT 4248  
Qy 4351 GATGAAAGCAGATGCCATGTGACACACTACCTGAAAGAGTTCCCAAGTGTCCCACAAT 4410  
Db 4249 GATGAAAGCAGATGCCATGTGACACACTACCTGAAAGAGTTCCCAAGTGTCCCACAAT 4308  
Qy 4411 ATTGCTTTTCTGATGTTTCACTGATGCACTAGTGCACATTTGACATGGAATGACCTGACCT 4470  
Db 4309 ATTGCTTTTCTGATGTTTCACTGATGCACTAGTGCACATTTGACATGGAATGACCTGACCT 4368  
Qy 4471 ATCCTTGGCTACTTTTCAAAAATTAACAAATTTACCACTCAACTGCTGCTCAAAAATGCAAA 4530  
Db 4369 ATCCTTGGCTACTTTTCAAAAATTAACAAATTTACCACTCAACTGCTGCTCAAAAATGCAAA 4428

Qy 4531 GAATGGGAATCCGAAGAATGTGTGAAATATCAAAAAATTTCAATACCTCTATGAAGCTCAC 4590  
Db 4429 GAATGGGAATCCGAAGAATGTGTGAAATATCAAAAAATTTCAATACCTCTATGAAGCTCAC 4488  
Qy 4591 TTAACCTGAAGAGACAGTATATGGAATTTAAAGAAATTTTAGATGGTATAGATTTCCAAGTGGCT 4650  
Db 4489 TTAACCTGAAGAGACAGTATATGGAATTTAAAGAAATTTTAGATGGTATAGATTTCCAAGTGGCT 4548  
Qy 4651 GCCAGCACCAATGCTGGCTATGGCAATGCTTCAAACTGGATTTTCTACAAAACTCTGCTCT 4710  
Db 4549 GCCAGCACCAATGCTGGCTATGGCAATGCTTCAAACTGGATTTTCTACAAAACTCTGCTCT 4608  
Qy 4711 GGGCTTCCAGATGGTCTCTCTGAAATGTTTCAATGATAGTAGCAACATCACCTTTTAGCATC 4770  
Db 4609 GGGCTTCCAGATGGTCTCTCTGAAATGTTTCAATGATAGTAGCAACATCACCTTTTAGCATC 4668  
Qy 4771 AGCATAAAGCTGGAGTGAACCTCTGTCATTTACTTGGACCAACATGTTTATCTGATGATGTC 4830  
Db 4669 AGCATAAAGCTGGAGTGAACCTCTGTCATTTACTTGGACCAACATGTTTATCTGATGATGTC 4728  
Qy 4831 AAATCGGTAGATAATGATGAATTTTAAATATATCTTCAATCAAGTCAAAATGAAGAAATATAA 4890  
Db 4729 AAATCGGTAGATAATGATGAATTTTAAATATATCTTCAATCAAGTCAAAATGAAGAAATATAA 4788  
Qy 4891 ACCATAGAAATTTAAAGATTTAGAAATTTTCAAGGATTTCTGTAGTGAATCACTGCAATTT 4950  
Db 4789 ACCATAGAAATTTAAAGATTTAGAAATTTTCAAGGATTTCTGTAGTGAATCACTGCAATTT 4848  
Qy 4951 ACTGGGAACATTTAGTGTCTGCATATGTAAGAGGAGTCAAGTCTGCTGCTGCTGCTGCTGCT 5010  
Db 4849 ACTGGGAACATTTAGTGTCTGCATATGTAAGAGGAGTCAAGTCTGCTGCTGCTGCTGCTGCT 4908  
Qy 5011 ACTTTAGATCAGCCCAAGGACCACTTAAACATGACATTTTCAAGAGATACACAGAT 5070  
Db 4909 ACTTTAGATCAGCCCAAGGACCACTTAAACATGACATTTTCAAGAGATACACAGAT 4968  
Qy 5071 GAAGTTTCAAAAATTTCAATTTAAAGTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5130  
Db 4969 GAAGTTTCAAAAATTTCAATTTAAAGTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5028  
Qy 5131 GTATATCAGCTCTGGTTTACCGAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5190  
Db 5029 GTATATCAGCTCTGGTTTACCGAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5088  
Qy 5191 AGTATTATACAGAAAACCAACATTCGTCATTTGCAATGCTAGAGGATGCTTAAAGGTGGA 5250  
Db 5089 AGTATTATACAGAAAACCAACATTCGTCATTTGCAATGCTAGAGGATGCTTAAAGGTGGA 5148  
Qy 5251 CATACATACATATCAGTGTGTTTACGCAATGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 5310  
Db 5149 CATACATACATATCAGTGTGTTTACGCAATGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 5208  
Qy 5311 ATGAGAAATACCATGATATCAAGCTCCAGCAGCAACCAAAACCAAAACCAAAACCAAAACCAAAAC 5370  
Db 5209 ATGAGAAATACCATGATATCAAGCTCCAGCAGCAACCAAAACCAAAACCAAAACCAAAACCAAAAC 5268  
Qy 5371 TATGATGCCACAGGAAACCTGCTGTGACTTCAACAAACAAATTTACATCAGATGCAAGTCAAGTCA 5430  
Db 5269 TATGATGCCACAGGAAACCTGCTGTGACTTCAACAAACAAATTTACATCAGATGCAAGTCAAGTCA 5328  
Qy 5431 TGTACTACTACAGTGTGATCATGGAACCAATTAATAAATTTGTAACAAGTGTCTGCGACAGAAACA 5490  
Db 5329 TGTACTACTACAGTGTGATCATGGAACCAATTAATAAATTTGTAACAAGTGTCTGCGACAGAAACA 5388  
Qy 5491 GGAGCTCAGCATGATGGAATGTAACAAAGTGGTATGATGATGATGATGATGATGATGATGATGATGAT 5550  
Db 5389 GGAGCTCAGCATGATGGAATGTAACAAAGTGGTATGATGATGATGATGATGATGATGATGATGATGAT 5448  
Qy 5551 CCATATTTTACAAATGAAGGCTTTCTTAAACCTTCCATGTCAGAGAGGAAAGCAAAAGTTT 5610  
Db 5449 CCATATTTTACAAATGAAGGCTTTCTTAAACCTTCCATGTCAGAGAGGAAAGCAAAAGTTT 5508

QY 5611 AGTGGCAATGAAGAAATCTACATCATAGGTGCTGATAATGCGATCATGATTCCTGGCAAT 5670  
Db 5509 AGTGGCAATGAAGAAATCTACATCATAGGTGCTGATAATGCGATCATGATTCCTGGCAAT 5568  
QY 5671 GAAGACAAAATTTGCAATGGACCACTGGAACCAAAAGCAATACATTTAAATTTAGA 5730  
Db 5569 GAAGACAAAATTTGCAATGGACCACTGGAACCAAAAGCAATACATTTAAATTTAGA 5628  
QY 5731 GCTACAAAATTTATGCGGCAATTTACTGACTCTGATTTATTCGACCTGTTAAGACTTTA 5790  
Db 5629 GCTACAAAATTTATGCGGCAATTTACTGACTCTGATTTATTCGACCTGTTAAGACTTTA 5688  
QY 5791 GGGGAAGGACTTTTCAGAAAGAACCGTAGAGATCATTTCTTCGCTCACTTTGTGTATCTTT 5850  
Db 5689 GCGGAAGGACTTTTCAGAAAGAACCGTAGAGATCATTTCTTCGCTCACTTTGTGTATCTTT 5748  
QY 5851 TCAATAAATTTCTCTTGGAAACAGCTAATTTTTCGATTTGCAAGAAATTCGACAGAAAGCAAAA 5910  
Db 5749 TCAATAAATTTCTCTTGGAAACAGCTAATTTTTCGATTTGCAAGAAATTCGACAGAAAGCAAAA 5808  
QY 5911 GAAGGTGGCACATCTCTCTCAGGATGCGAGAAATTTATTCACACTAAATTTGAAGCTGGAT 5970  
Db 5809 GAAGGTGGCACATCTCTCTCAGGATGCGAGAAATTTATTCACACTAAATTTGAAGCTGGAT 5868  
QY 5971 CAGCTCATCAGTGGCAGACCTTGGAACTGAAGGACGAGAGATTAAACGCG----- 6020  
Db 5869 CAGCTCATCAGTGGCAGACCTTGGAACTGAAGGACGAGAGATTAAACGCGTTACTTACT 5928  
QY 6021 -----GCCAATAAGCAAGAAATCTCTCTCCTGCAACATGTTGAAGAGCTTT 6063  
Db 5929 TATAGAAAATCCATCAAGCCAAATAGCAAGAAATCTCTCTCCTGCAACATGTTGAAGAGCTTT 5988  
QY 6064 TGCACAAACCAACCTTAAGTTTCAAGAGAAATTTTCGGAATTTACCAAAATTTCTTTCAG 6123  
Db 5989 TGCACAAACCAACCTTAAGTTTCAAGAGAAATTTTCGGAATTTACCAAAATTTCTTTCAG 6048  
QY 6124 GATCTTTCTTCAACTGATGCTGATCTGCTTGGAAATAGAGCAAAACCGTTTCCCAAC 6183  
Db 6049 GATCTTTCTTCAACTGATGCTGATCTGCTTGGAAATAGAGCAAAACCGTTTCCCAAC 6108  
QY 6184 ATAAACCAATATAATTAATAACAGAGTAAAGCTGATGATGAGCTGATGTTTCCAGGT 6243  
Db 6109 ATAAACCAATATAATTAATAACAGAGTAAAGCTGATGATGAGCTGATGTTTCCAGGT 6165  
QY 6244 TCGGATTAATTAATGCGCAGCTATATTTCTGTTTATTTATGCTCCAAATGAATTTATGCT 6303  
Db 6166 TCGGATTAATTAATGCGCAGCTATATTTCTGTTTATTTATGCTCCAAATGAATTTATGCT 6225  
QY 6304 ACTCAAGGTCCACTACAGGAAACAGTTGGAGATTTTGGAGAAATGTTGGGAAACCCAGG 6363  
Db 6226 ACTCAAGGTCCACTACAGGAAACAGTTGGAGATTTTGGAGAAATGTTGGGAAACCCAGG 6285  
QY 6364 GCAAAACATTAATGTAATGTAACACAGTGTGTTTGAAGAAAGGACGAGATCGCATCAG 6423  
Db 6286 GCAAAACATTAATGTAATGTAACACAGTGTGTTTGAAGAAAGGACGAGATCGCATCAG 6345  
QY 6424 TATTGGCCAGAGACCAACAGCCAGTTACTGCTTTGGAGATATAGTGAATTAACAAGCTA 6483  
Db 6346 TATTGGCCAGAGACCAACAGCCAGTTACTGCTTTGGAGATATAGTGAATTAACAAGCTA 6405  
QY 6484 ATGGAGGATGTTCAAAATAGATTGGACTATCAGGATCTGAAATTTGAAGGATGGGAT 6543  
Db 6406 ATGGAGGATGTTCAAAATAGATTGGACTATCAGGATCTGAAATTTGAAGGATGGGAT 6465  
QY 6544 TGCATGACTGTTTCGACAGTGTAACTTTACTGCTGGCCAGAGATGGGGTTCTGAGAAC 6603  
Db 6466 TGCATGACTGTTTCGACAGTGTAACTTTACTGCTGGCCAGAGATGGGGTTCTGAGAAC 6525  
QY 6604 AGCGCCCTCTAAATTCATCTTTGGAAGTTGGTTTCGAGCAAGCAGGCGCATGACACCA 6663  
Db 6526 AGCGCCCTCTAAATTCATCTTTGGAAGTTGGTTTCGAGCAAGCAGGCGCATGACACCA 6585  
QY 6664 CCTATGATTTGTTCACTGCGAGTGTGAGTTGGAAGAACTGGAGTTTTTTATTGCTCTGGAC 6723

Db 6586 CCTATGATTTGTTCACTGTAGTCTGGAGTTGGAAGAACTGGAGTTTTTTATTGCTCTGGAC 6645  
QY 6724 CATTTAAACAACATATAAATGACCATGATTTTGTGGATATATATGAGTAGTAGCTGAA 6783  
Db 6646 CATTTAAACAACATATAAATGACCATGATTTTGTGGATATATATGAGTAGTAGCTGAA 6705  
QY 6784 CTGAGAAGTGAAGAAATGTGATGTGCAAGAACTGGCAGACATATATCTTTTACACACAG 6843  
Db 6706 CTGAGAAGTGAAGAAATGTGATGTGCAAGAACTGGCAGACATATATCTTTTACACACAG 6765  
QY 6844 TGCATTTGCGATCTCTTATCAAAATGAAGGAAGTAATCAAGCCCATCTGTTTGTGTAACAT 6903  
Db 6766 TGCATTTGCGATCTCTTATCAAAATGAAGGAAGTAATCAAGCCCATCTGTTTGTGTAACAT 6825  
QY 6904 TCAGCACTTCAGAAAGATGACTCTTTGGACGCCATGGAAGGTGATGTCAGCTTGAATGG 6963  
Db 6826 TCAGCACTTCAGAAAGATGACTCTTTGGACGCCATGGAAGGTGATGTCAGCTTGAATGG 6885  
QY 6964 GAAGAAACCCACTATGTAA 6981  
Db 6886 GAAGAAACCCACTATGTAA 6903  
RESULT 7  
ABT06281  
ID ABT06281 standard; cDNA; 6994 BP.  
XX  
AC ABT06281;  
XX  
DT 24-OCT-2002 (first entry)  
XX  
DE Human NOV2a coding sequence.  
XX  
KW Human; NOVX; autoimmune disease; cancer; infection; inflammatory disease; storage disorder; muscle disorder; neurodegenerative disorder; nontropic; developmental defect; neuroprotective; antiparkinsonian; hypotensive; hypertensive; haemostatic; cardiac; antitumoral; dermatological; immunosuppressive; antiinflammatory; virucide; antibacterial; anti-HIV; antiparasitic; antiallergic; antidiabetic; immunomodulator; antipsoriatic; vulnery; anorectic; antidiabetic; antitumor; cerebroprotective; anticonvulsant; nephrotoxic; kerolytic; antitumor; antidiabetic; immunomodulator; antipsoriatic; antifertility; antitumor; antidiabetic; anticonvulsant; anticonvulsant; tranquilizer; analgesic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200257450-A2.  
XX  
PD 25-JUL-2002.  
XX  
PF 29-NOV-2001; 2001WO-US048922.  
XX  
PR 29-NOV-2000; 2000US-0253834P.  
PR 30-NOV-2000; 2000US-0250926P.  
PR 25-JAN-2001; 2001US-0264180P.  
PR 20-AUG-2001; 2001US-0313656P.  
PR 05-OCT-2001; 2001US-0327456P.  
PR 28-NOV-2001; 2001US-00327456.  
XX  
(CURA-) CURAGEN CORP.  
XX  
PI Edinger S, Macdougall JR, Millet J, Ellerman K, Stone DJ;  
PI Gerlach V, Grosse WM, Alsobrook JP, Lepley DM, Rieser D, Burgess CE;  
PI Casman SJ, Spytek KA, Boldog FL, Li L, Padigar M, Mishra V;  
PI Patturajan M, Shenoy S, Rastelli L, Tchernev VT, Vernet CM;  
PI Zerhusen BD, Malyankar UM, Guo X, Miller CE, Gangolli EA;  
XX  
DR WPI: 2002-590741/63.  
XX P-PSDB; AA018736.  
PT Novel isolated polypeptide, designated NOVX, useful for treating or preventing in NOVX-associated disorders e.g. cardiomyopathy,

PT atherosclerosis, diabetes, cancer, allergy, asthma, Crohn's disease.  
XX Claim 9; Page 23-25; 353pp; English.

CC The present invention provides the protein and coding sequences of  
CC several novel human proteins, designated NOVX. These can be used in the  
CC treatment of, amongst others, cancers, autoimmune diseases, infections,  
CC inflammatory diseases, storage disorders, muscle disorders,  
CC neurodegenerative diseases and developmental defects. The present  
CC sequence is a coding sequence of the invention

XX SQ Sequence 6994 BP; 2330 A; 1388 C; 1320 G; 1956 T; 0 U; 0 Other;

Query Match 90.8%; Score 6453; DB 6; Length 6994;

Best Local Similarity 96.3%; Pred. No. 0;

Matches 6770; Conservative 0; Mismatches 215; Indels 48; Gaps 14;

QY 76 TGAATCTACTGGCTGAAATATGTAATAAGATGATTTCTTATCATTTTCTTTACTT 135

DB 1 TGAATCTACTGGCTGAAATATGTAATAAGATGATTTCTTATCATTTTCTTTACTT 60

QY 136 TTTATTGGGACTTCAGAGACACAGGTTGATGTTTCCAATGTCGTTCTGGTACTAGGTAC 195

DB 61 TTTATTGGGACTTCAGAGACACAGGTTGATGTTTCCAATGTCGTTCTGGTACTAGGTAC 120

QY 196 GATATAACCATCTCTCAATTTCTAACAATACACCTCACCTGTTACTAGATAGTGACA 255

DB 121 GATATAACCATCTCTCAATTTCTAACAATACACCTCACCTGTTACTAGATAGTGAGCT 180

QY 256 CCAATGTAACAAACCGAGGCTCCAGTCTTCTAGCGGGGAAAGTCCGATCTGCT 315

DB 181 TCTAATG-----AACAGGCTCCAGTCTTCTAGCGGGGAAAGTCCGATCTGCT 234

QY 316 GGGATCTCTCTGTTGGATATACACCACTTAATCCAAATGGAAGGATTAATCTTACAT 375

DB 235 GGGATCTCTCTGTTGGATATACACCACTTAATCCAAATGGAAGGATTAATCTTACAT 294

QY 376 GTCAATATATAGGAAGTTTCTCGTGGATGCAACAGTATATATACAAAGTCAAG 435

DB 295 GTCAATATATAGGAAGTTTCTCGTGGATGCAACAGTATATATACAAAGTCAAG 354

QY 436 CAGACAGTCTGGAAGTTCTTCTTACTTAATCTTAATCTTAACTGGACCAATATGAATTAAG 495

DB 355 CAGACAGTCTGGAAGTTCTTCTTACTTAATCTTAACTGGACCAATATGAATTAAG 414

QY 496 GTTGTGCTGCAAAACAGTCTGCAATGGAGTGTGTAGTGCATCTTCTTCTCCAACT 555

DB 415 GTAGCTGCTGAAACAGTCTGCAATGGAGTGTGTAGTGCATCTTCTTCTCCAACT 474

QY 556 GCAGAAAGTCTCCAGGAAAGTGGTGAATCTCACAGTTGAGCGCTACACGCTTCAGCA 615

DB 475 GCAGAAAGTCTCCAGGAAAGTGGTGAATCTCACAGTTGAGCGCTACACGCTTCAGCA 532

QY 616 GTTAAGCTGATTTGTTATTTACTTCGGACCAATGCGCAATTAACAGCTTCAAGAT 675

DB 533 G-TAAGCTGATGGTA--TACCTCGGCAACCAAA--AAAAAATTAACAGCTTCAAGAT 588

QY 676 AGTGTCAACATGCCAGAGTGGATAGTAGTGAAGATGTCTCAATCAGAGTAGAGGAC 735

DB 589 AGTGTCAACATGCCAGAGTGGATAGTAGTGAAGATGTCTCAATCAGAGTAGAGGAC 648

QY 736 ATTTTGAATGGGAAATTTGCC--AGAATGCAATGAGATAGTGAATCTTTTATGGAGT 792

DB 649 ATTTTGAATGGTCTCCCTTCTCTTGGCACTGCAACAGAGATAGTGAATCTTTTATGGAGT 708

QY 793 ACAGCAGCCCTCTCCAACTTGTGTAGGTTACACCTCCATCGGCTACCAATTC 852

DB 709 ACAGCAGCCCTCTCTCCAACTTGTGTAGGTTACACCTCCATCGGCTACCAATTC 768

QY 853 TCAAGCAGCTTGACAGAGATGATCAGCTGTGTGGAAAGAGCTATCAGTTTGTGA 912

DB 769 TCAAGCAGCTTGACAGAGATGATCAGCTGTGTGGAAAGAGCTATCAGTTTGTGA 825

QY 913 GTGACACACTTGAGACCTTATATACACATATCTTTTTGAAAGTTTCAGCTGCTCAACTGAA 972

DB 826 GTGACACACTTGAGACCTTATATACACATATCTTTTTGAAAGTTTCAGCTGCTCAACTGAA 885

QY 973 GCAGGTTATATGATAGTACGATTGTGAGAACACACAGAAATCAGTGCCTGAAGACCA 1032

DB 886 GCAGGTTATATGATAGTACGATTGTGAGAACACACAGAAATCAGTGCCTGAAGACCA 945

QY 1033 CAAACTGCGTAAACAGGCAACATCACAGGAAAGTCTCTTTCAATTTTATGGGACCA 1092

DB 946 CAAACTGCGTAAACAGGCAACATCACAGGAAAGTCTCTTTCAATTTTATGGGACCA 1005

QY 1093 ACTATAGTAACAGGGAATTTAGTTATAGAGTTGAATATATAGGACAT---CAGTCCG 1149

DB 1006 ACTATAGTAACAGGGAATTTAGTTATAGAGTTGAATATATAGGACATCAGCAGGTCG 1065

QY 1150 ATTTTGGATAACAGCAAAAAGACCTCAAGTTTGGCAATTCACAACTTAACACCATTTACA 1209

DB 1066 ATTTTGGATAACAGCAAAAAGACCTCAAGTTTGGCAATTCACAACTTAACACCATTTACA 1125

QY 1210 ATGTATGATGTCTATATTGCGCTGAAACAGTGCAGGAGCTGGGCCCAAGTCAATATTT 1269

DB 1126 ATGTATGATGTCTATATTGCGCTGAAACAGTGCAGGAGCTGGGCCCAAGTCAATATTT 1185

QY 1270 TCGATATTCACTCCACAGATGTTCCAGGGGAGTGTGTGATTTTACAATTGACAGAGGTA 1329

DB 1186 TCGATATTCACTCCACAGATGTTCCAGGGGAGTGTGTGATTTTACAATTGACAGAGGTA 1245

QY 1330 GAATCCACCAAGTAGTAGAATTTCTGGAAGAACACACAGCAACCAATGGAATTTATTAAC 1389

DB 1246 GAATCCACCAAGTAGTAGAATTTCTGGAAGAACACACAGCAACCAATGGAATTTATTAAC 1305

QY 1390 CAATACCGAGTGAAGTGTAGTTCCAGAGACAGGAATTAATTTTGGAAATACTTTGCTC 1449

DB 1306 CAATACCGAGTGAAGTGTAGTTCCAGAGACAGGAATTAATTTTGGAAATACTTTGCTC 1365

QY 1450 ACTGGAATTAAGTATTAATTAATGACCCCATGGCTCCAGAAATTTGTGAACATAGTAGAG 1509

DB 1366 ACTGGAATTAAGTATTAATTAATGACCCCATGGCTCCAGAAATTTGTGAACATAGTAGAG 1422

QY 1510 CCAATGGTAGGATTAATGAGGGTTCCAGCAGAGATGTCGCTGACCTTCACTCACTTGCT 1569

DB 1423 CCAATGGTAGGATTAATGAGGGTTCCAGCAGAGATGTCGCTGACCTTCACTCACTTGCT 1482

QY 1570 ACATTTATATATAACAGCCATCCAGATAAAACCTTCTCCAGAGGAATGAGCTGAAGAC 1629

DB 1483 ACATTTATATATAACAGCCATCCAGATAAAACCTTCTCCAGAGGAATGAGCTGAAGAC 1542

QY 1630 CAGACTTCAACAGTTGTAACTACAGGAATCAGTATATTAATGACATTTGAGCTGAACAG 1689

DB 1543 CAGACTTCAACAGTTGTAACTACAGGAATCAGTATATTAATGACATTTGAGCTGAACAG 1602

QY 1690 CTGCTTATGTTATCAGGAGCTTGTACCTTTCACTGAGCAGATGATGATGCTGCT 1749

DB 1603 CTGACTTATGTTCTTATCAGATTAAGGAGATTTGGGCTGAGACATTTGGGGTTTTCTAGA 1662

QY 1750 TTCCACCATCATGGGAGAGGACCAACAGTGTCTCAGTGTGTAGGACAGCTCAGCAGGTG 1809

DB 1663 TATACATCATG-----TCACTCTGCAAGCAGGAGCAATTTGACTTCCCGAGG 1710

QY 1810 CCAAGCTCCATTTAAATTAATAAACTATAAAATAATTTAGTGTCTTCACTATTTTGTGTATAT 1869

DB 1711 CTTTGTGAGCCCAAAATTTTCAAGATTTACATGTTTACCATTAACAGAGATTTTATACAC 1770

QY 1870 TGGGATCCTCCAGATATCCCAATGGAAAAATAAATCACTACTATACGATTTTATGCAATGGA 1929

DB 1771 TGGGATCCTCCAGATCCTGTAATTTTTCATCATTAATTAATTTTATGATGTTGAA 1830

QY 1930 TTTGGATACAAAACAGAGCTTCCAGATAACTACATAGATAACAGCTTTCTCTAATACAGGG 1989

DB 1831 AACCAATCCAGAGATTAATTTTATAGGACATTTAAACAGATTTGTCTCTCTTATAGGG 1890

QY 1990 TTAAGAAATATACAAAATACAAAATAGAGTGGAGCTCAACCCACGATGGAGAAAGT 2049

1891 TTAAGAAATACACAAAATACAAAATGAGAGTGGCAGCCTCAACCCACGTTGGAGAACT 1950  
1951 TCTTTGCTGAGAAAATGACATCTTTGTGAGAACTTCAAGAGATGAAACCGGAATCATCA 2010  
2011 CCTCAAGATGTCGAGTAAATGATGTTACCGCAGATGAAATAGGTTGAAGTGGTCACCA 2169  
2070 CCTCAAGATGTCGAGTAAATGATGTTACCGCAGATGAAATAGGTTGAAGTGGTCACCA 2070  
2170 CCGGAAAGCCCAATGGGATCATTTATGCTTATGAAGTGTCTATATAAAAATATAGATCT 2229  
2071 CCGGAAAGCCCAATGGGATCATTTATGCTTATGAAGTGTCTATATAAAAATATAGATCT 2130  
2230 TTTATATATGAGAAACATCAACACAGACATATAATTAAGAACTTAAGACCTCACCC 2289  
2131 TTTATATATGAGAAACATCAACACAGACATATAATTAAGAACTTAAGACCTCACCC 2190  
2290 CTCTATAACATTTCTGTAAGTCTTACACAGATTTGGTCAATGGCAATCAGGTATCTTCT 2349  
2191 CTCTATAACATTTCTGTAAGTCTTACACAGATTTGGTCAATGGCAATCAGGTATCTTCT 2250  
2350 TTAATCTCTGTAAGGACTTCGGAGACTGTGCTGTAGTGCCAGTGTGACAGAAATATCACTTAC 2409  
2251 TTAATCTCTGTAAGGACTTCGGAGACTGTGCTGTAGTGCCAGTGTGACAGAAATATCACTTAC 2310  
2410 AAAAAATATTTCTTGGAGAGATTTGAGCTATCATCTCTTCCCTCCCAAGTGTCCCAATGGA 2469  
2311 AAAAAATATTTCTTGGAGAGATTTGAGCTATCATCTCTTCCCTCCCAAGTGTCCCAATGGA 2370  
2470 ATCATATAAAAAATATACAAATTTATCTCAAGAGAGTAAATGGAATCAGGAAAGAACTATA 2529  
2371 ATCATATAAAAAATATACAAATTTATCTCAAGAGAGTAAATGGAATCAGGAAAGAACTATA 2430  
2530 AATAACAACCTCTTTAAACCCAAACATTAAGTACTGAAGAAATATACCCAAATATATCATTT 2589  
2431 AATAACAACCTCTTTAAACCCAAACATTAAGTCTGAAGAAATATACCCAAATATATCATTT 2490  
2590 GAGGTGCTGCTAGTATACATGAAGGTGAAGGATTCGGAGTGTCCCAATAGTATACTG 2649  
2491 GAGGTGCTGCTAGTATACATGAAGGTGAAGGATTCGGAGTGTCCCAATAGTATACTG 2550  
2650 ACGGAGGAGATGCTCTGATCTCTCCCTCAAGACTTCTCTGTAAGAACTGTTGCTGTGT 2709  
2551 ACGGAGGAGATGCTCTGATCTCTCCCTCAAGACTTCTCTGTAAGAACTGTTGCTGTGT 2610  
2710 GTCACGGTGAAGTGTGTCAGGCAACCAACCCCTGGAGCCAAATGGAATATCTCTTTATTAC 2769  
2611 GTCACGGTGAAGTGTGTCAGGCAACCAACCCCTGGAGCCAAATGGAATATCTCTTTATTAC 2670  
2770 ACGTTTATGTCGGAAATAGATCATCATTTAAATAATTAATGTCATGAAACATCATTTG 2829  
2671 ACGTTTATGTCGG--AGATCATCATTTAAATAATTAATGTCATGAAACATCATTTG 2727  
2830 GAGTTATCAGATTTGAGTATATATGTTGAATACAGTCTTATGTCAGCTAGCAGCCAGA 2889  
2728 GAGTTATCAGATTTGAGTATATATGTTGAATACAGTCTTATGTTAAACAGCTAGCAGCCAGA 2787  
2890 TTTTGTGATGGGAAAAACAGGAACATATCATTTAGCTTTCAACACAGAGGAGCACC 2949  
2788 TTTTGTGATGGGAAAAACAGGAACATATCATTTAGCTTTCAACACAGAGG--ACCA 2844  
2950 AGCGATCTCCCAAGATGTTTATATGCAACCTCAGTCTTTCATCAATATCTTTTC 3009  
2845 AGCGATCTCCCAAGATGTTTATATGCAACCTCAGTCTTTCATCAATATCTTTTC 2904  
3010 TGGACACCTCTTCAAAACCTTAATGGATTTATACAAATTTACTCTGTTTATTACAGAAAT 3069  
2905 TGGACACCTCTTCAAAACCTTAATGGATTTATACAAATTTACTCTGTTTATTACAGAAAT 2964  
3070 ACTTCAGGTACTTTTATGCAAAATTTTACATCTCCATGAACCTAACCAATGACTTTGCAAT 3129

2965 ACTTCAGGTACTTTTATGCAAAATTTTACATCTCCATGAAGTAAACCAATGACTTTGCAAT 3024  
3130 ATGACTGTATCCCAATATATAGATAAACTGACAAATATTCAGCTACTATATACATTTTGGTTA 3189  
3025 ATGACTGTATCCCAATATATAGATAAACTGACAAATATTCAGCTACTATATACATTTTGGTTA 3084  
3190 ACAGCAAGTACTTCAGTTTGGAAATGGGAATAAAAGCAGTGACATCATTTGAAGTATACACA 3249  
3085 ACAGCAAGTACTTCAGTTTGGAAATGGGAATAAAAGCAGTGACATCATTTGAAGTATACACA 3144  
3250 GATCAAGACATACCTGGAAGGTTTGTGGAAACCTGACATTAAGAACTTCTTCTCAACT 3309  
3145 GATCAAGACATACCTGGAAGGTTTGTGGAAACCTGACATTAAGAACTTCTTCTCAACT 3204  
3310 GCAATAAATGTAAGCTGGGTCCACCGCTCAACCAACGGTCTAGTCTTCTTACTATGTT 3369  
3205 GCAATAAATGTAAGCTGGGTCCACCGCTCAACCAACGGTCTAGTCTTCTTACTATGTT 3264  
3370 TCACTGATCTTACAGCAGACTCTCTGCCATGTGAGACCACTCTTGTGTTACATATGAGAGA 3429  
3265 TCACTGATCTTACAGCAGACTCTCTGCCATGTGAGACCACTCTTGTGTTACATATGAGAGA 3324  
3430 AGCATATATTTTGAATACTGGAATAATACATGATATATATATAAATAATCTCATCA 3489  
3325 AGCATATATTTTGAATACTGGAATAATACATGATATATATATAAATAATCTCATCA 3384  
3490 ACAGAAAGGATCTCTGATACCTATCTGCCAGCTATACATCAAGACTCAAGAAAGAT 3549  
3385 ACAGAAAGGATCTCTGATACCTATCTGCCAGCTATACATCAAGACTCAAGAAAGAT 3444  
3550 GTTCCCAAGAACTTCCCAATAATCAACACTTTTAAACCTTTTCTTCTTCTTCTTCTTCT 3609  
3445 ATCCCAAGAACTTCCCAATAATCAACACTTTTAAACCTTTTCTTCTTCTTCTTCTTCT 3504  
3610 TTTATATGAGGATCCCCAGTAAAGCAAAATGCTGCAATTAATGATTAATGATTAATCTTTA 3669  
3505 TTTATATGAGGATCCCCAGTAAAGCAAAATGCTGCAATTAATGATTAATGATTAATCTTTA 3564  
3670 CAAGGACCAATGAAATTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3729  
3565 CAAGGACCAATGAAATTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3624  
3730 CTTTCAACATTTACATATATATAGCTTTTGTGCGCAGAACTAGAAAAGGACTTGT 3789  
3625 CTTTCAACATTTACATATATATAGCTTTTGTGCGCAGAACTAGAAAAGGACTTGT 3684  
3790 CTTTCAAGTATTTCTTTCTTTTACAGATGATGCTGCGCTGAGCACTGCAAAAT 3849  
3685 CTTTCAAGTATTTCTTTCTTTTACAGATGATGCTGCGCTGAGCACTGCAAAAT 3744  
3850 TTTGACTTTTAACTCACTGACTTTCAGACTTTTGTATGCTGAAATGGAGCCCAAGTCTCTT 3909  
3745 TTTGACTTTTAACTCACTGACTTTCAGACTTTTGTATGCTGAAATGGAGCCCAAGTCTCTT 3804  
3910 CCAGTGTGTTTGTGTTTAAAGTATATAGTTTAAATTTTCAATGAACTGAAATCTGACATA 3969  
3805 CCAGTGTGTTTGTGTTTAAAGTATATAGTTTAAATTTTCAATGAACTGAAATCTGACATA 3864  
3970 TATATAGAAATATATCAGGATTTTAACTGAAAGCAAACTTGTGAGCTGGAACAGCTC 4029  
3865 TATATAGAAATATATCAGGATTTTAACTGAAAGCAAACTTGTGAGCTGGAACAGCTC 3924  
4030 AGCAGCTCTCTATCGTGTATCTGTTTCCAAAGTTTGGAAATGGCAATCAATTTAGT 4089  
3925 AGCAGCTCTCTATCGTGTATCTGTTTCCAAAGTTTGGAAATGGCAATCAATTTAGT 3984  
4090 AATGTAGTAAATTTCAACACCCCAAGAACTCAGTTTCCAGATGTCGTGAGAAATGCAAGTGC 4149  
3985 AATGTAGTAAATTTCAACACCCCAAGAACTCAGTTTCCAGATGTCGTGAGAAATGCAAGTGC 4044  
4150 ATGGGAACTAGCTGGAGTCAAGTTTGTAGTAAATGGGATCCACCCAAAAGGCAATGGA 4209  
4045 ATGGCAACTAGCTGGAGTCAAGTTTGTAGTAAATGGGATCCACCCAAAAGGCAATGGA 4104



QY 4210 ATAAATACGCGATATATGTTAAAGGAAATTTCTACAAAAGTTTCTCCCAAGAT 4269  
DB 4105 ATAAATACGCGATATATGTTAAAGGAAATTTCTACAAAAGTTTCTCCCAAGAT 4164  
QY 4270 CACATGTACACTTTTCAAAAGCTTTCTGCAATACCTCATATGCTCTTTAAAGTAAGGCT 4329  
DB 4165 CACATGTACACTTTTCAAAAGCTTTCTGCAATACCTCATATGCTCTTTAAAGTAAGGCT 4224  
QY 4330 TCAACCTCAGCTGGTGAAGGTGATGAAGACATGCGCATGTCAGACACACTACCTGAAACA 4389  
DB 4225 TCAACCTCAGCTGGTGAAGGTGATGAAGACATGCGCATGTCAGACACACTACCTGAAACA 4284  
QY 4390 GTTCCAGGTGTTCCCAAAATATGCTTTTCTGATGTTCAAGTCAACTAGTCAAAATG 4449  
DB 4285 GTTCCAGGTGTTCCCAAAATATGCTTTTCTGATGTTCAAGTCAACTAGTCAAAATG 4344  
QY 4450 ACATGATAGACTGACACTATCTTGGCTACTTTTCAAAATTTACAAAATTTACCACTCA 4509  
DB 4345 ACATGATAGACTGACACTATCTTGGCTACTTTTCAAAATTTACAAAATTTACCACTCA 4404  
QY 4510 CTTCTGCTCAAAAATGCAAAAGTGGAAATCCGAAGATGCTGCAATATCAAAAATTT 4569  
DB 4405 CTTCTGCTCAAAAATGCAAAAGTGGAAATCCGAAGATGCTGCAATATCAAAAATTT 4464  
QY 4570 CAATACCTCTATGAAGCTCACTTAACTGAAGACAGTATATGATTTAAAGAAATTTAGA 4629  
DB 4465 CAATACCTCTATGAAGCTCACTTAACTGAAGACAGTATATGATTTAAAGAAATTTAGA 4524  
QY 4630 TGGTATAGATTCGAAGTGGCTGCGACGACCAATGCTGGCTATGGCAATGCTTCAAACTGG 4689  
DB 4525 TGGTATAGATTCGAAGTGGCTGCGACGACCAATGCTGGCTATGGCAATGCTTCAAACTGG 4584  
QY 4690 ATTTCTACAAAATCTGCTGCTGCTCCAGATGCTCCCTGCAAAATGTTCAATGATGA 4749  
DB 4585 ATTTCTACAAAATCTGCTGCTGCTCCAGATGCTCCCTGCAAAATGTTCAATGATGA 4644  
QY 4750 GCAACATCACTTTTAGCATCAGCATAAAGCTGGAAGTGAACCTGCTGCTCAATTTACTGCA 4809  
DB 4645 GCAACATCACTTTTAGCATCAGCATAAAGCTGGAAGTGAACCTGCTGCTCAATTTACTGCA 4704  
QY 4810 ACATGTTATCTGATGTCATCAATCGGTAGATTAATGATGAATTTAATATATCTTCTATC 4869  
DB 4705 ACATGTTATCTGATGTCATCAATCGGTAGATTAATGATGAATTTAATATATCTTCTATC 4764  
QY 4870 AAGTCAAAATGAAGAAATTAACCATAGAAATTAAGATTTAGAAATTTTCAAGGTAT 4929  
DB 4765 AAGTCAAAATGAAGAAATTAACCATAGAAATTTAGAAATTTTCAAGGTAT 4824  
QY 4930 TCTGTAGTATCATCTGCAATTTACTGGGAACATTTAGTGTGTCATPATGTAGAAGGAAAGTCA 4989  
DB 4825 TCTGTAGTATCATCTGCAATTTACTGGGAACATTTAGTGTGTCATPATGTAGAAGGAAAGTCA 4884  
QY 4990 AGTGTGAAATGATGTTACTTCTAGATCAGCCCAAGGACCCACCTAACACATG 5049  
DB 4885 AGTGTGAAATGATGTTACTTCTAGATCAGCCCAAGGACCCACCTAACACATG 4944  
QY 5050 ACATTTTCAAGAGATACAGATGAAGTTTCAAAATTTCAATTAAGCTTTCTTCTCTCTCT 5109  
DB 4945 ACATTTTCAAGAGATACAGATGAAGTTTCAAAATTTCAATTAAGCTTTCTTCTCTCTCTCT 5004  
QY 5110 CAACTTAATGGAATATCCAAGTATATCAAGCTCTGTTTACCGAGAAGATGATCTACT 5169  
DB 5005 CAACTTAATGGAATATCCAAGTATATCAAGCTCTGTTTACCGAGAAGATGATCTACT 5064  
QY 5170 GCTGTCCAGATTCACAACTCAGTATATACAGAAACCAACATTCGTTCATGCAATG 5229  
DB 5065 GCTGTCCAGATTCACAACTCAGTATATACAGAAACCAACATTCGTTCATGCAATG 5124  
QY 5230 CTAGAAGGCTAAAGGTGACATACATCAATATCAGTCTTTTACCGAGTCAATAGTCT 5289  
DB 5125 CTAGAAGGCTAAAGGTGACATACATCAATATCAGTCTTTTACCGAGTCAATAGTCT 5184

QY 5290 GGTGCGAGTCCAAAGGTTCCGATGAGAAATAACCATGGATATCAAGCTCCAGCACACCA 5349  
DB 5185 GGTGCGAGTCCAAAGGTTCCGATGAGAAATAACCATGGATATCAAGCTCCAGCACACCA 5244  
QY 5350 AAAACCCAAACCAACCCCTATTTATGATGCCACAGGAAACTGCTGTGACTTCAACAACA 5409  
DB 5245 AAAACCCAAACCAACCCCTATTTATGATGCCACAGGAAACTGCTGTGACTTCAACAACA 5304  
QY 5410 ATTACAATCAGATGCCAATATTTTACTACAGTGATGATCATGGACCAATAAAAAATGTA 5469  
DB 5305 ATTACAATCAGATGCCAATATTTTACTACAGTGATGATCATGGACCAATAAAAAATGTA 5364  
QY 5470 CAAAGTCTTGGACAGAAACAGAGCTCAGCATGATGGAATGTAACAAGTGGTATGAT 5529  
DB 5365 CAAAGTCTTGGACAGAAACAGAGCTCAGCATGATGGAATGTAACAAGTGGTATGAT 5424  
QY 5530 GCATATTTTAAATAAGAGGCGCATATTTTACAATGAAGGCTTTTCTCAACCTCCATGT 5589  
DB 5425 GCATATTTTAAATAAGAGGCGCATATTTTACAATGAAGGCTTTTCTCAACCTCCATGT 5484  
QY 5590 ACAGAAGGAAAGACAAAGTTTAGTGGCAATGAAGAAATCTACATCATAGGTGCTGATAAT 5649  
DB 5485 ACAGAAGGAAAGACAAAGTTTAGTGGCAATGAAGAAATCTACATCATAGGTGCTGATAAT 5544  
QY 5650 GCATGATGATTTCTGGCAATGAAGACAAATTTGCAATGGACCACTGAAACCAAAAAG 5709  
DB 5545 GCATGATGATTTCTGGCAATGAAGACAAATTTGCAATGGACCACTGAAACCAAAAAG 5604  
QY 5710 CAATCTATTTAAATTTAGAGCTACAAATATTTATGGGCAATTTTACTGACTCTGATAT 5769  
DB 5605 CAATCTATTTAAATTTAGAGCTACAAATATTTATGGGCAATTTTACTGACTCTGATAT 5664  
QY 5770 TCTGACCTCTGTTAAGACTTTTAGGGGAAGGACTTTTCAAGAAAGAACCGTAGAGATCATTTCT 5829  
DB 5665 TCTGACCTCTGTTAAGACTTTTAGGGGAAGGACTTTTCAAGAAAGAACCGTAGAGATCATTTCT 5724  
QY 5830 TCGGTCACTTTGTGATCTCTTCAATATTTCTCTTGGAAACAGCTATTTTGTGATGCA 5889  
DB 5725 TCGGTCACTTTGTGATCTCTTCAATATTTCTCTTGGAAACAGCTATTTTGTGATGCA 5784  
QY 5890 AGAATTCGACAGACAGAAAGAGTGGCACATCTCTCAGAGATCAGAAATTTAT 5949  
DB 5785 AGAATTCGACAGACAGAAAGAGTGGCACATCTCTCAGAGATCAGAAATTTAT 5844  
QY 5950 GACACTAAATTTGAAGCTGGATCAGCTCATCAGTGGCAGACCTGGAACCTGAAGGACGAG 6009  
DB 5845 GACACTAAATTTGAAGCTGGATCAGCTCATCAGTGGCAGACCTGGAACCTGAAGGACGAG 5904  
QY 6010 AGATTAACGCGGCCAATAAGCAAGAAATCTCTTCTGCAACATGTTGAAGAGCTTTGCACA 6069  
DB 5905 AGATTAACGCGGCCAATAAGCAAGAAATCTCTTCTGCAACATGTTGAAGAGCTTTGCACA 5964  
QY 6070 AACACACCTAAAGTTTCAAGAGAAATTTTCCGAATTTACCAAAATTTCTTCAGGATCTT 6129  
DB 5965 AACACACCTAAAGTTTCAAGAGAAATTTTCCGAATTTACCAAAATTTCTTCAGGATCTT 6024  
QY 6130 TCTTCAACTGATGCTGATCTGCTTGGAAATAGCAAAAAACCGTTTCCCAACATAAAA 6189  
DB 6025 TCTTCAACTGATGCTGATCTGCTTGGAAATAGCAAAAAACCGTTTCCCAACATAAAA 6084  
QY 6190 CCAATATAATAATAACAGAGTAAGCTGATGATGACGCTAGTGTTCAGGTTCCGAT 6249  
DB 6085 CCAATATAATAATAACAGAGTAAGCTGATGATGACGCTAGTGTTCAGGTTCCGAT 6141  
QY 6250 TATATTAATGCCAGCTATATTTCTGGTATTTATGTCCAAAATGAAATTTTCTACTCAA 6309  
DB 6142 TATATTAATGCCAGCTATATTTCTGGTATTTATGTCCAAAATGAAATTTTCTACTCAA 6201  
QY 6310 GGTCCACTACCAAGAAACAGTTCGAGATTTTTCGAGAAATGGTGTGGGAAACCCAGGGCAAAA 6369  
DB 6202 GGTCCACTACCAAGAAACAGTTCGAGATTTTTCGAGAAATGGTGTGGGAAACCCAGGGCAAAA 6261  
QY 6370 ACATTAGTAATGCTAACACAGTGTGTTTGAAGAAAGGACCGATCAGATGCTATGTTGG 6429



Db 6262 ACATTAGTAATGCTAACACAGTCTTTTGAARAAGGACGGATCAGATGCCATCAGTATGG 6321  
QY 6430 CCAGAGGACAAACAGCGAGTTACTGTCTTTGGAGATATAGTGATTAACAAAGCTAATGGAG 6489  
Db 6322 CCAGAGGACAAACAGCGAGTTACTGTCTTTGGAGATATAGTGATTAACAAAGCTAATGGAG 6381  
QY 6490 GATGTTCAATAGATTGGAGTATCAGGGATCTGAAAATTTGAAGGATGGGATTTGCATG 6549  
Db 6382 GATGTTCAATAGATTGGAGTATCAGGGATCTGAAAATTTGAAGGATGGGATTTGCATG 6441  
QY 6550 ACTGTTCCGACAGTGAATCTTACTGCTGCCAGACATGGGGTTCTCGAGAACACGCGC 6609  
Db 6442 ACTGTTCCGACAGTGAATCTTACTGCTGCCAGACATGGGGTTCTCGAGAACACGCGC 6501  
QY 6610 CTTCTAATTAATCACTTTGTGAAGTTGGTTTCGAGCAAGCAGGGCACAATGACACCACTATG 6669  
Db 6502 CTTCTAATTAATCACTTTGTGAAGTTGGTTTCGAGCAAGCAGGGCACAATGACACCACTATG 6561  
QY 6670 ATTGTTCACTGCGAGTCTGGAGTTGGAAGAACTGGAGTTTATTGCTCTGGACCAATTGA 6729  
Db 6562 ATTGTTCACTGCGAGTCTGGAGTTGGAAGAACTGGAGTTTATTGCTCTGGACCAATTGA 6621  
QY 6730 ACACACATATAAATGACCATGATTTTGTGGATATATATGACTAGTCTGAATCTGAGA 6789  
Db 6622 ACACACATATAAATGACCATGATTTTGTGGATATATATGACTAGTCTGAATCTGAGA 6681  
QY 6790 AGTGAAGAATGCGATGTTGCGAGAATCTGGCAAGTATATCTTTTTTACACCACTGCAAT 6849  
Db 6682 AGTGAAGAATGCGATGTTGCGAGAATCTGGCAAGTATATCTTTTTTACACCACTGCAAT 6741  
QY 6850 CTGGATCTCTTATCAATTAAGGAAGTAAATCAGCCCATCTGTTTCTTAACTATTTCAGCA 6909  
Db 6742 CTGGATCTCTTATCAATTAAGGAAGTAAATCAGCCCATCTGTTTCTTAACTATTTCAGCA 6801  
QY 6910 CTTTCAAGATGACTCTTTGGACGCGCATGGAA ---GGTGAATGTCAGCTTGAATGGGAA 6966  
Db 6802 CTTTCAAGATGACTCTTTGGACGCGCATGGAGTGGTGAATGTCAGCTTGAATGGGAA 6861  
QY 6967 GAAACCACTATGTAATATTTAGACCAAGGATACAAATTTGGAAGAGATTTTAAATCCCA 7026  
Db 6862 GAAACCACTATGTAATATTTAGACCAAGGATACAAATTTGGAAGAGATTTTAAATCCCA 6921  
QY 7027 GGGGCCAAAGTTACCCCTCATCTTCCGAATTTGAATGTGCAACCTTAAAGAAATATCT 7086  
Db 6922 GGGGCCAAAGTTACCCCTCATCTTCCGAATTTGAATGTGCAACCTTAAAGAAATATCT 6981  
QY 7087 ATGCTTCTCTCAC 7099  
Db 6982 ATGCTTCTCTCAC 6994

RESULT 8

ADH41618

ID ADH41618 standard; DNA; 6994 BP.

XX

AC ADH41618;

XX

DT 25-MAR-2004 (first entry)

XX

DE Novel human nucleic acid NOV15d.

XX

ds; gene; cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic; antidiabetic; immunosuppressive; anti-HIV; neuroprotective; neurotrophic; antiparkinsonian; antiaesthetic; antinfertility; cardiomyopathy; atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS; multiple sclerosis; graft-versus-host disease; Alzheimer's disease; Parkinson's disease; asthma; fertility disorder; chromosome mapping; tissue typing; preventive medicine; pharmacogenomic; vaccine.

XX

OS Homo sapiens.

XX

XX WO2003102159-A2.

XX 11-DEC-2003.  
PD  
XX  
PF 04-JUN-2003; 2003WO-US017573.  
XX  
PR 04-JUN-2002; 2002US-0385490P.  
PR 04-JUN-2002; 2002US-0385615P.  
PR 04-JUN-2002; 2002US-0385755P.  
PR 05-JUN-2002; 2002US-0386041P.  
PR 06-JUN-2002; 2002US-0386355P.  
PR 06-JUN-2002; 2002US-0386447P.  
PR 06-JUN-2002; 2002US-0386459P.  
PR 06-JUN-2002; 2002US-0386465P.  
PR 06-JUN-2002; 2002US-0386864P.  
PR 07-JUN-2002; 2002US-0386701P.  
PR 07-JUN-2002; 2002US-0386796P.  
PR 07-JUN-2002; 2002US-0386931P.  
PR 07-JUN-2002; 2002US-0387078P.  
PR 07-JUN-2002; 2002US-0387081P.  
PR 07-JUN-2002; 2002US-0387083P.  
PR 10-JUN-2002; 2002US-0387429P.  
PR 10-JUN-2002; 2002US-0387540P.  
PR 10-JUN-2002; 2002US-0387866P.  
PR 11-JUN-2002; 2002US-0387606P.  
PR 11-JUN-2002; 2002US-0387610P.  
PR 11-JUN-2002; 2002US-0387659P.  
PR 11-JUN-2002; 2002US-0387668P.  
PR 11-JUN-2002; 2002US-0387696P.  
PR 11-JUN-2002; 2002US-0387859P.  
PR 12-JUN-2002; 2002US-0387934P.  
PR 12-JUN-2002; 2002US-0387960P.  
PR 12-JUN-2002; 2002US-0388022P.  
PR 12-JUN-2002; 2002US-0388096P.  
PR 12-JUN-2002; 2002US-0388432P.  
PR 12-JUN-2002; 2002US-0388479P.  
PR 13-JUN-2002; 2002US-0389123P.  
PR 14-JUN-2002; 2002US-0389120P.  
PR 14-JUN-2002; 2002US-0389146P.  
PR 17-JUN-2002; 2002US-0389742P.  
PR 18-JUN-2002; 2002US-0389604P.  
PR 18-JUN-2002; 2002US-0389884P.  
PR 19-JUN-2002; 2002US-0390006P.  
PR 19-JUN-2002; 2002US-0390144P.  
PR 19-JUN-2002; 2002US-0390209P.  
PR 25-JUN-2002; 2002US-0391726P.  
PR 06-AUG-2002; 2002US-0401628P.  
PR 09-AUG-2002; 2002US-0402268P.  
PR 12-AUG-2002; 2002US-0402822P.  
PR 13-AUG-2002; 2002US-0403458P.  
PR 15-AUG-2002; 2002US-0403617P.  
PR 15-AUG-2002; 2002US-0403732P.  
PR 26-AUG-2002; 2002US-0406182P.  
PR 12-SEP-2002; 2002US-0410085P.  
PR 13-SEP-2002; 2002US-0410505P.  
PR 23-SEP-2002; 2002US-0412955P.  
PR 30-SEP-2002; 2002US-0415195P.  
PR 23-OCT-2002; 2002US-0420627P.  
PR 23-OCT-2002; 2002US-0420718P.  
PR 24-OCT-2002; 2002US-0420852P.  
PR 31-OCT-2002; 2002US-0422750P.  
PR 01-NOV-2002; 2002US-0423095P.  
PR 05-NOV-2002; 2002US-0423748P.

(CURA-) CURAGEN CORP.

XX  
XX  
PI Alsbrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;  
PI Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;  
PI Ellerman K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;  
PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;  
PI Khrantsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;  
PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CB;  
PI Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli I;

PI Rieger DX, Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G;  
PI Szytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernhet CAM, Voss E2;  
PI Wolenc AR, Zhong H, Zhong H;  
XX  
DR WPI: 2004-053467/05.  
DR P-PSDB; ADH41619.  
XX  
XX New NOVX polypeptides and nucleic acid molecules useful for preventing or  
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in  
PT pharmacogenomics.  
XX  
XX Claim 20; SEQ ID NO 171; 1503pp; English.  
XX  
XX The invention relates to 566 new isolated human polypeptides and their  
CC encoding genes, sequences that are at least 95% identical to these or  
CC sequences comprising one or more conservative substitutions in these or  
CC polypeptide, polynucleotide and antibodies against the polypeptides are  
CC useful in diagnosing, treating or preventing NOVX-associated disorders,  
CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,  
CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,  
CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.  
CC The nucleic acids are further used as hybridization probes, in chromosome  
CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The  
CC polypeptides are also useful as vaccines. This sequence represents an  
CC example of the nucleic acid sequence of the invention.  
XX  
SQ Sequence 6994 BP; 2330 A; 1388 C; 1320 G; 1956 T; 0 U; 0 Other;

Query Match 90.8%; Score 6453; DB 12; Length 6994;  
Best Local Similarity 96.3%; Pred. No. 0;  
Matches 6770; Conservative 0; Mismatches 215; Indels 48; Gaps 14;

QY 76 TGATTTCTACTGGCTGAAATGTAATAAGATGATTTCTTATCAATTTCTTTTACTT 135  
DB 1 TGATTTCTACTGGCTGAAATGTAATAAGATGATTTCTTATCAATTTCTTTTACTT 60

QY 136 TTTATTTGGGACTTCAGAGACACAGGTTGATGTTTCCAAATGTCGTTCTGTAAGTAC 195  
DB 61 TTTATTTGGGACTTCAGAGACACAGGTTAGATGTTTCCAAATGTCGTTCTGTAAGTAC 120

QY 196 GATATAACCATCTCTTCAATTTCTACAACATACACCTCCTGTTACTAGATAGTAC 255  
DB 121 GATATAACCATCTCTTCAATTTCTACAACATACACCTCCTGTTACTAGATAGGCT 180

QY 256 CCAATGTAAACAAAACAGGGCTCCAGTCTTCTAGCCGGGAAAGAGTCGATCTGCT 315  
DB 181 TCTATG-----AACAGGGCTCCAGTCTTCTAGCCGGGAAAGAGTCGATCTGCT 234

QY 316 GGGATCTTCTGTTGGGAATACACCACTTAATCCAAATGGAGGATATATCTTACATT 375  
DB 235 GGGATCTTCTGTTGGGAATACACCACTTAATCCAAATGGAGGATATATCTTACATT 294

QY 376 GTCAATATATAAGAAAGTTGTCGGTGATGCAACAGTATATACACAGTCAGATCAAG 435  
DB 295 GTCAATATATAAGAAAGTTGTCGGTGATGCAACAGTATATACACAGTCAGATCAAG 354

QY 436 CCAGACAGTCTGGAAGTCTTCTTACTTAATCTTAATCTTGGAAACACATATGAATTAAG 495  
DB 355 CCAGACAGTCTGGAAGTCTTCTTACTTAATCTTAATCTTGGAAACACATATGAATTAAG 414

QY 496 GTTGTGCTGTAACACAGTCTGTCATTTGGAGTGTGTTAGTGATCCATTCTCTTCCAACT 555  
DB 415 GTTGTGCTGTAACACAGTCTGTCATTTGGAGTGTGTTAGTGATCCATTCTCTTCCAACT 474

QY 556 GCAGAAAGTCTCCAGAAAAGTGGTGAATCTCACAGTTGAGGCTTACAAAGCTTCAGCA 615  
DB 475 GCAGAAAGTCTCCAGAAAAGTGGTGAATTTTACAGGTCAGGCTGTC--CGTTTCAGCA 532

QY 616 GTTAAGCTGATTTGGTATTTACTCTGGCAACCAATGGCAAAATACAGCTTCAAGATT 675  
DB 533 G-TAAGCTGATTTGGTATTTACTCTGGCAACCAATGGCAAAATACAGCTTCAAGATT 588

QY 676 AGTGTCAAACATGCCAGAAAGTGGGATAGTAGTGAAAGATGTCTCAATCAGAGTAGAGGAC 735  
DB 589 AGTGTCAAGCATAAACAGAAAGTGGGATAGTAGTGAAAGATGTCTCAATCAGAGTAGAGTGC 648

QY 736 ATTTTCAGCTGGGAAATTTGCC---AGAAATCAATGAGATAGTGAATCTTTTATCGAGT 792  
DB 649 ATTTTAAGTGTCTTCCCTTCTTGGCACTGCAACGAGATAGTGAATCTTTTATGGAGT 708

QY 793 ACAGCCAGCCCTTCTTCCAAACCTTGTAGAGTTTACACCTCCATCGGATACACACATTTCA 852  
DB 709 ACAGCCAGCCCTTCTTCCAAACCTTGTAGAGTTTACACCTCCATCGGATACACACATTTCA 768

QY 853 TCAGCAGCTTGACACAGAAATGAGATCAGCTCTGTGTGAAAGAGCTATCAGTTTGTGA 912  
DB 769 TCAAGCAGCTTGACACAGAAATGAGATCAGCTCTGT---GAAAGAGCTATCAGTTTGTGA 825

QY 913 GTGACACACTTCAGACCTTATACACATATCTTTTGAAGTTTTCAGCTGCTCAACACTGAA 972  
DB 826 GTGACACACTTCAGACCTTATACACATATCTTTTGAAGTTTTCAGCTGCTCAACACTGAA 885

QY 973 GCAGGTTATATTGATAGTACGATTTGTGAGAACACCAAGATCAGTGCTGGAAGCACCACCA 1032  
DB 886 GCAGGTTATATTGATAGTACGATTTGTGAGAACACCAAGATCAGTGCTGGAAGCACCACCA 945

QY 1033 CAAACTGCTGAACAGGCAACATCACAGGAAATGCTCTTTTCAATTTTATGGGACCCACCA 1092  
DB 946 CAAACTGCTGAACAGGCAACATCACAGGAAATGCTCTTTTCAATTTTATGGGACCCACCA 1005

QY 1093 ACTNTAGTAAACAGGGAATTTAGTTATAGAGTTGAAATATATGGACAT---CAGTCCG 1149  
DB 1006 ACTNTAGTAAACAGGGAATTTAGTTATAGAGTTGAAATATATGGACATCAGCAGTCCG 1065

QY 1150 ATTTTGGATAAACAGCAACAAAGACCTCAAGTTTGCATTCACTAACTTAACACCATTTACA 1209  
DB 1066 ATTTTGGATAAACAGCAACAAAGACCTCAAGTTTGCATTCACTAACTTAACACCATTTACA 1125

QY 1210 ATGTATGATGCTATATTTCGGCTGAAACAGTGCAGGAGCTGGGCCCAAGTCAATATTT 1269  
DB 1126 ATGTATGATGCTATATTTCGGCTGAAACAGTGCAGGAGCTGGGCCCAAGTCAATATTT 1185

QY 1270 TCAGTATTTACCTCCACAGATGTTCCAGGGGAGTGTGTTGATTTTAACTTCAGAGGTA 1329  
DB 1186 TCAGTATTTACCTCCACAGATGTTCCAGGGGAGTGTGTTGATTTTAACTTCAGAGGTA 1345

QY 1330 GAATCCACGCAAGTAAAGATTTACTTTGGAAGAAACACGACCAACCAATGGAATTTAATAC 1389  
DB 1246 GAATCCACGCAAGTAAAGATTTACTTTGGAAGAAACACGACCAACCAATGGAATTTAATAC 1305

QY 1390 CAATACCGAGTCAAAAGTGTCTAGTTCCAGAGACAGGAATTAATTTGGAAATACTTTGCTC 1449  
DB 1306 CAATACCGAGTCAAAAGTGTCTAGTTCCAGAGACAGGAATTAATTTGGAAATACTTTGCTC 1365

QY 1450 ACTGGAAATAGTATATATAATGACCCATGGCTCCAGAAATTTGTGAACATAGTAGAG 1509  
DB 1366 ACTGGAAATAGTATATAATGACCCATGGCTCCAGAAATTTGTGAACATAGTAGAG 1422

QY 1510 CCAATGGTAGGATATATAGGAGTTTTCAGCAGAGATGTCGCTGACCTTCACTTCACTTGTCT 1569  
DB 1423 CCAATGGTAGGATATATAGGAGTTTTCAGCAGAGATGTCGCTGACCTTCACTTGTCT 1482

QY 1570 ACATTTATATATAACAGCCATCCAGATATAAACTTTTCTGCAAGGAATAGAGCTGAAGAC 1629  
DB 1483 ACATTTATATATAACAGCCATCCAGATATAAACTTTTCTGCAAGGAATAGAGCTGAAGAC 1542

QY 1630 CAGACTTCCAGTTGTAATCAAGGAATCAGTATATATCTGACATTTGACGCTGAACAG 1689  
DB 1543 CAGACTTCCAGTTGTAATCAAGGAATCAGTATATATCTGACATTTGACGCTGAACAG 1602

QY 1690 CTGCTTATGTTTATCAGGAGCTTGTACCTTTTCACTGAGCAGATGATTTAGTGTATCTGCT 1749  
DB 1603 CTGACTTATGTTTCTTATCAGATTAAGGAGATTTTGGGCTGAGACATGGGGTTTCTAGA 1662

QY 1750 TTCACCATCATGGGAGAGGACCAACCAACAGTTCTCAGTGTGTTAGGACACGTCAGCAAGTG 1809

1663 TATACATCATG-----TCATCTGCAAGCAGGACAAATTTTGACTTCCCAGGC 1710  
1810 CCAAGCTCCATTTAAATTAATAAATAATATAGTCTCTTCATCTATTTTGGTTATAT 1869  
1711 CTTTGTGACGCCCAAAATTTGAGGTACACATGTTTACCATAAACAGAAATATTTTACAC 1770  
1870 TGGGATCCTCCAGAAATATCCCAATGGAATAAATCACTATATACGATTTATCCATGGAA 1929  
1771 TGGGATCCTCCAGATCCTGTATTTTTCATCATTTACCTTATCATTTTGGATGTGAA 1830  
1930 TTGGATCAAAACAGAGCAATTCAGATAAATACCATAGATAAAGAGCTTCTCATACAGGG 1989  
1831 AACCANTCCAGNGTATTTTAAAGGACATTAACAGTTTGTCTCTTCTTATAGG 1890  
1990 TTAAGAAATACACAAAATACAAAATGAGAGTGGCAGCTCAACCCACGATGGAGAAAT 2049  
1891 TTAAGAAATACACAAAATACAAAATGAGAGTGGCAGCTCAACCCACGTTGGAGAAAT 1950  
2050 TCTTTGTCTGAAGAAATGACATCTTTGTGAGAACTTCAGAGATGAACCGGAATCATCA 2109  
1951 TCTTTGTCTGAAGAAATGACATCTTTGTGAGAACTTCAGAGATGAACCGGAATCATCA 2010  
2110 CTTCAAGATGTCGAAGTAAATGATGTTACCGCAGATGAATAAGGTTGAAGTGGTCACCA 2169  
2011 CTTCAAGATGTCGAAGTAAATGATGTTACCGCAGATGAATAAGGTTGAAGTGGTCACCA 2070  
2170 CCGGAAAGCCCAATCGGATCAATTTATGCTTTATGAAGTGTCTATATAAAAATATAGATCT 2229  
2071 CCGGAAAGCCCAATCGGATCAATTTATGCTTTATGAAGTGTCTATATAAAAATATAGATCT 2130  
2230 TTATATATGAAGAACATCAACACAGACATATATTAAGNACTTAAGACTCACACC 2289  
2131 TTATATATGAAGAACATCAACACAGACATATATTAAGNACTTAAGACTCACACC 2190  
2290 CTTATAAATTTCTGTGAAGTCTTACACAGATTTGGTTCATGGCAATCAGGTATCTTCT 2349  
2191 CTTATAAATTTCTGTGAAGTCTTACACAGATTTGGTTCATGGCAATCAGGTATCTTCT 2250  
2350 TTACTCTCTGAAGGACTTCGGAGATGAGCTATCAATCTTCCCCCAAGTAGTCCCAATGA 2409  
2251 TTACTCTCTGAAGGACTTCGGAGATGAGCTATCAATCTTCCCCCAAGTAGTCCCAATGA 2310  
2410 AAAAAATTTCTTCTCGAGAGTTGAGCTATCAATCTTCCCCCAAGTAGTCCCAATGA 2469  
2311 AAAAAATTTCTTCTCGAGAGTTGAGCTATCAATCTTCCCCCAAGTAGTCCCAATGA 2370  
2470 ATCATAAAAAATATACAAATTTATCTCAAGAGAGTAAATGGAATGAGGAAGAACTATA 2529  
2371 ATCATACAAAATATACAAATTTATCTCAAGAGAGTAAATGGAATGAGGAAGAACTATA 2430  
2530 AATACAACTCTTTAAACCCAAAACNTTAAAGTACTCAAGAAATATATCCCAATATATCAT 2589  
2431 AATACAACTCTTTAAACCCAAAACNTTAAAGTCTCAAGAAATATATCCCAATATATCAT 2490  
2590 GAGGTCTGCTAGTACACTGAAGGTGAAGGAGTTGGAGTCTCCCATAGTATACG 2649  
2491 GAGGTCTGCTAGTACACTGAAGGTGAAGGAGTTGGAGTCTCCCATAGTATACG 2550  
2650 ACGGAGGAGATGCTCTGATTTCTCCCTCAAGACTTCTCTGTAAGAAAGTGTCTGCT 2709  
2551 ACGGAGGAGATGCTCTGATTTCTCCCTCAAGACTTCTCTGTAAGAAAGTGTCTGCT 2610  
2710 GTCACGGTGAAGTGTGATGGCAACCAACCCCTGGAGCCAAATGAATATCTTTTATATAC 2769  
2611 GTCACGGTGAAGTGTGATGGCAACCAACCCCTGGAGCCAAATGAATATCTTTTATATAC 2670  
2770 ACAGTTTATCTCGGAATAGATCATCATTAAGAACTATTAATGCTCACTGAACATCATTTG 2829  
2671 ACAGTTTATGCTCGG---AGATCATCATTAAGAACTATTAATGCTCACTGAACATCATTTG 2727  
2830 GAGTTATCAGATTTGGATTAATATGTTGAATACAGTGTCTTATGTAAACAGCTAGCACCA 2889

2728 GAGTTATCAGATTTGGATTAATAATGTTGAATACAGTGTCTTATGTAAACAGCTAGCACCA 2787  
2890 TTTTGTGTGATGGAAACAGGAGCAATATATCTAGCTTTCAAAACACAGAGGAGCACCA 2949  
2788 TTTTGTGTGATGGAAACAGGAGCAATATATCTAGCTTTCAAAACACAGAGG---ACCA 2844  
2950 AGCGATCCTCCCAAGATGTTTATATGCAAACTCAGTTCCTCATCAATAATTTCTTTTC 3009  
2845 AGCGATCCTCCCAAGATGTTTATATGCAAACTCAGTTCCTCATCAATAATTTCTTTTC 2904  
3010 TGGACACCTCTCTTCAAAACCTTAATGGGATATACAAATATCTCTGTTTATTCAGAAAT 3069  
2905 TGGACACCTCTCTTCAAAACCTTAATGGGATATACAAATATCTCTGTTTATTCAGAAAT 2964  
3070 ACTTCAGGTACTTTTATGCGAATTTTACACTCCATGAACTAACCAATGACCTTTCGACAT 3129  
2965 ACTTCAGGTACTTTTATGCGAATTTTACACTCCATGAACTAACCAATGACCTTTCGACAT 3024  
3130 ATGACTGTATCCCAATATATAGATAAACTGACAAATATTCAGCTACTATATACATTTTGGTTA 3189  
3025 ATGACTGTATCCCAATATATAGATAAACTGACAAATATTCAGCTACTATATACATTTTGGTTA 3084  
3190 ACAGCAAGTACTTCAGTTGGAAATGGGAAATAAAGCAGTGACATCATTTGAAGTATACACA 3249  
3085 ACAGCAAGTACTTCAGTTGGAAATGGGAAATAAAGCAGTGACATCATTTGAAGTATACACA 3144  
3250 GATCAAGACATACCTGAAAGGTTTCTTGGAAACCTGACTTACGAATCCATTTTCGCAACT 3309  
3145 GATCAAGACATACCTGAAAGGTTTCTTGGAAACCTGACTTACGAATCCATTTTCGCAACT 3204  
3310 GCAATAAATGTAAGCTGGGTCCCAACCGCTCAACCAAAACGGTCTAGTCTTCTTACTATGTT 3369  
3205 GCAATAAATGTAAGCTGGGTCCCAACCGCTCAACCAAAACGGTCTAGTCTTCTTACTATGTT 3264  
3370 TCACTGATCTTACAGCAGACTCTCGCCATGTGAGACCACTCTTGTGTACATATGAGAGA 3429  
3265 TCACTGATCTTACAGCAGACTCTCGCCATGTGAGACCACTCTTGTGTACATATGAGAGA 3324  
3430 AGCATATATTTGATAATCTGGAAATAATACACTGATATATATTAATAAATACTCCATCA 3489  
3325 AGCATATATTTGATAATCTGGAAATAATACACTGATATATATTAATAAATACTCCATCA 3384  
3490 ACAGAAAAGGGATTTCTGATACCTATCTGCCCCAGCTATACATCAAGCTGAAGAAGAT 3549  
3385 ACAGAAAAGGGATTTCTGATACCTATCTGCCCCAGCTATACATCAAGCTGAAGAAGAT 3444  
3550 GTCCCAAGAACTTCACCAATAATCAACACTTTTAAACCTTTCTTACCTCAGTTCTC 3609  
3445 ATCCCAAGAACTTCACCAATAATCAACACTTTTAAACCTTTCTTACCTCAGTTCTC 3504  
3610 TTATCATGGGATCCCCAGTAAAGCCAAATGGTGAATAATAAGTTATGATTTAACTTTA 3669  
3505 TTATCATGGGATCCCCAGTAAAGCCAAATGGTGAATAATAAGTTATGATTTAACTTTA 3564  
3670 CAAGAACCAAAATGAAAATTTATTTCTTCACTTCTGATAATTTACATAATTTTGAAGAG 3729  
3565 CAAGAACCAAAATGAAAATTTATTTCTTCACTTCTGATAATTTACATAATTTTGAAGAG 3624  
3730 CTTTCAACCAATTTACATATATAGCTTTTGTGCGCGCAAGAACTAGAAAAGGACTTGGT 3789  
3625 CTTTCAACCAATTTACATATATAGCTTTTGTGCGCGCAAGAACTAGAAAAGGACTTGGT 3684  
3790 CTTTCCAGTATTTCTTTTCTTTTACACAGATGAGTCCGCTGAGCTCCACCAAAAT 3849  
3685 CTTTCCAGTATTTCTTTTCTTTTACACAGATGAGTCCGCTGAGCTCCACCAAAAT 3744  
3850 TTGACTTTTAAATCAACTGTACTTTCAGACTTTTGTATGGCTGAAATGGAGCCCAAGTCTCTT 3909  
3745 TTGACTTTTAAATCAACTGTACTTTCAGACTTTTGTATGGCTGAAATGGAGCCCAAGTCTCTT 3804  
3910 CAGGTGGTATTTGTTAAAGTATATAGTTTTTAAATTTCAATGAACATGAACATGACTATA 3969  
3805 CAGGTGGTATTTGTTAAAGTATATAGTTTTTAAATTTCAATGAACATGAACATGACTATA 3864



Db 6025 TCTTCACTGATGCTGATCTGCTTGGATAGAGCAAAAACCGCTTCCCAACATATAA 6084  
QY 6190 CCATATAATAATAAACAAGATAAGCTAGTAGCTAGCGCTAGTGTTCAGGTTCCGAT 6249  
Db 6085 CCAT---ATAATAATAACAAGATAAGCTAGTAGCTAGCGCTAGTGTTCAGGTTCCGAT 6141  
QY 6250 TATATTATGCCAGCTATATTTCTGGTATTTATCTCCAAATGAATTTATTCCTACTCAA 6309  
Db 6142 TATATTATGCCAGCTATATTTCTGGTATTTATCTCCAAATGAATTTATTCCTACTCAA 6201  
QY 6310 GGTCCACTACCAAGAACAGTTCGAGATTTTCGAGAATGGTGGGAACACAGGCGCAAAA 6369  
Db 6202 GGTCCACTACCAAGAACAGTTCGAGATTTTCGAGATGGTGGGAACACAGGCAAAA 6261  
QY 6370 ACATTAGTAATCTAACACAGTGTTTTGAAGAACGCGATCAGATGCCATCAGTATTGG 6429  
Db 6262 ACATTAGTAATCTAACACAGTGTTTTGAAGAACGCGATCAGATGCCATCAGTATTGG 6321  
QY 6430 CCAGAGGACAAACAGCCAGTTACTGCTTTGGAGATATAGTCAATACAAAGCTAATGGAG 6489  
Db 6322 CCAGAGGACAAACAGCCAGTTACTGCTTTGGAGATATAGTCAATACAAAGCTAATGGAG 6381  
QY 6490 GATGTTCAATAGATTGCACTATCAGGATCTGAAATTTGAAGGATGGGATTCGATG 6549  
Db 6382 GATGTTCAATAGATTGCACTATCAGGATCTGAAATTTGAAGGATGGGATTCGATG 6441  
QY 6550 ACTGTTGCAAGTGAATCTTACTGCTGCGCAGAGCATGCGGTTCTTGAGAACAGCGCC 6609  
Db 6442 ACTGTTGCAAGTGAATCTTACTGCTGCGCAGAGCATGCGGTTCTTGAGAACAGCGCC 6501  
QY 6610 CCTCTAATTCATTTGTGAAGTTGTTGAGCAAGCGGCACATGACACCACTATG 6669  
Db 6502 CCTCTAATTCATTTGTGAAGTTGTTGAGCAAGCGGCACATGACACCACTATG 6561  
QY 6670 ATTGTTCACTGCAAGTCTGGAGTTGGAAGAACTGGAGTTTTTATTTGCTCTGACCAATTA 6729  
Db 6562 ATTGTTCACTGCAAGTCTGGAGTTGGAAGAACTGGAGTTTTTATTTGCTCTGACCAATTA 6621  
QY 6730 ACACAACATATAAATGACCATGATTTTGTGGATATATATGAGTACTAGTGAATCTGAGA 6789  
Db 6622 ACACAACATATAAATGACCATGATTTTGTGGATATATATGAGTACTAGTGAATCTGAGA 6681  
QY 6790 AGTGAAGAAATGTGCATGTTGAGAACTCTGCGACAGTATATCTTTTACACCAAGTGCATT 6849  
Db 6682 AGTGAAGAAATGTGCATGTTGAGAACTCTGCGACAGTATATCTTTTACACCAAGTGCATT 6741  
QY 6850 CTGGATCTCTTATCAATAAGGGAAGTAACTAGCCCATCTGTTTTTGTAACTATTACGCA 6909  
Db 6742 CTGGATCTCTTATCAATAAGGGAAGTAACTAGCCCATCTGTTTTTGTAACTATTACGCA 6801  
QY 6910 CTTGAGAAGATGCACTCTTTGAGCGCCATGGAA---GGTGAATGTTGAGCTTGAATGGGAA 6966  
Db 6802 CTTGAGAAGATGCACTCTTTGAGCGCCATGGAAAGGTGGTGAATGTTGAGCTTGAATGGGAA 6861  
QY 6967 GAAACCACTATCTAATATTTCAGACCAAGGATCAATTTGGAAGAGATTTTAAATCCCA 7026  
Db 6862 GAAACCACTATCTAATATTTCAGACCAAGGATCAATTTGGAAGAGATTTTAAATCCCA 6921  
QY 7027 GGGGGCAAAAGTTACCCCTCATCTTCCGAAATGCAAACTTAAAGAAATATCT 7086  
Db 6922 GGGGGCAAAAGTTACCCCTCATCTTCCGAAATGCAAACTTAAAGAAATATCT 6981  
QY 7087 ATGCTTCTCTCAC 7099  
Db 6982 ATGCTTCTCTCAC 6994

RESULT 9  
ADH41612  
ID ADH41612 standard; DNA; 6881 BP.  
XX  
AC ADH41612;

XX 25-MAR-2004 (first entry)  
DT Novel human nucleic acid NOV15a.  
DE ds; gene; cardiovascular; antiarteriosclerotic; hypotensive; cytostatic;  
KW anorectic; antidiabetic; immunosuppressive; anti-HIV; neuroprotective;  
KW neotropic; antiparkinsonian; antiasthmatic; antiinfertility;  
KW cardiomyopathy; atherosclerosis; hypertension; cancer; obesity; diabetes;  
KW AIDS; multiple sclerosis; graft-versus-host disease; Alzheimer's disease;  
KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;  
KW tissue typing; preventive medicine; pharmacogenomic; vaccine.  
XX Homo sapiens.  
OS  
XX WO2003102159-A2.  
FN 11-DEC-2003.  
PD  
XX 04-JUN-2003; 2003WO-US017573.  
PF 04-JUN-2002; 2002US-0385490P.  
XX 04-JUN-2002; 2002US-0385615P.  
PR 04-JUN-2002; 2002US-0385755P.  
PR 05-JUN-2002; 2002US-0386041P.  
PR 06-JUN-2002; 2002US-0386355P.  
PR 06-JUN-2002; 2002US-0386357P.  
PR 06-JUN-2002; 2002US-0386447P.  
PR 06-JUN-2002; 2002US-0386459P.  
PR 06-JUN-2002; 2002US-0386465P.  
PR 06-JUN-2002; 2002US-0386864P.  
PR 07-JUN-2002; 2002US-0386701P.  
PR 07-JUN-2002; 2002US-0386796P.  
PR 07-JUN-2002; 2002US-0386931P.  
PR 07-JUN-2002; 2002US-0387078P.  
PR 07-JUN-2002; 2002US-0387081P.  
PR 10-JUN-2002; 2002US-0387429P.  
PR 10-JUN-2002; 2002US-0387540P.  
PR 10-JUN-2002; 2002US-0387866P.  
PR 11-JUN-2002; 2002US-0387606P.  
PR 11-JUN-2002; 2002US-0387610P.  
PR 11-JUN-2002; 2002US-0387659P.  
PR 11-JUN-2002; 2002US-0387668P.  
PR 11-JUN-2002; 2002US-0387696P.  
PR 11-JUN-2002; 2002US-0387853P.  
PR 12-JUN-2002; 2002US-0387934P.  
PR 12-JUN-2002; 2002US-0388022P.  
PR 12-JUN-2002; 2002US-0388096P.  
PR 12-JUN-2002; 2002US-0388432P.  
PR 12-JUN-2002; 2002US-0388479P.  
PR 13-JUN-2002; 2002US-0389123P.  
PR 14-JUN-2002; 2002US-0389120P.  
PR 14-JUN-2002; 2002US-0389146P.  
PR 17-JUN-2002; 2002US-0389742P.  
PR 18-JUN-2002; 2002US-0389604P.  
PR 18-JUN-2002; 2002US-0389884P.  
PR 19-JUN-2002; 2002US-0390006P.  
PR 19-JUN-2002; 2002US-0390144P.  
PR 19-JUN-2002; 2002US-0390209P.  
PR 25-JUN-2002; 2002US-0391726P.  
PR 06-AUG-2002; 2002US-0401628P.  
PR 09-AUG-2002; 2002US-0402268P.  
PR 12-AUG-2002; 2002US-0402822P.  
PR 13-AUG-2002; 2002US-0403458P.  
PR 15-AUG-2002; 2002US-0403617P.  
PR 15-AUG-2002; 2002US-0403732P.  
PR 26-AUG-2002; 2002US-0406182P.  
PR 12-SEP-2002; 2002US-0410085P.  
PR 13-SEP-2002; 2002US-0410503P.  
PR 23-SEP-2002; 2002US-0412955P.  
PR 30-SEP-2002; 2002US-0415195P.

PR 23-OCT-2002; 2002US-0420627P.  
 PR 23-OCT-2002; 2002US-0420718P.  
 PR 24-OCT-2002; 2002US-0420852P.  
 PR 31-OCT-2002; 2002US-0422750P.  
 PR 01-NOV-2002; 2002US-0423095P.  
 PR 05-NOV-2002; 2002US-0423748P.  
 XX (CURA-) CURAGEN CORP.  
 PI Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL,  
 PI Burgess CE, Casman SJ, Catterton E, Dhanabai M, Edinger SR;  
 PI Ellerman K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;  
 PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;  
 PI Khramtsov NV, Larochelele WJ, Li L, Liang H, Low K, Macdougall JR;  
 PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;  
 PI Millet I, Padigara M, Peyman JA, Qian X, Rastelli L;  
 PI Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G;  
 PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;  
 PI Wolenc AR, Zhong M, Zhong H;  
 XX  
 DR WPI; 2004-053467/05.  
 DR P-PSDB; ADH41613.  
 XX  
 PT New NOVX polypeptides and nucleic acid molecules useful for preventing or  
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
 PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in  
 PT pharmacogenomics.  
 XX  
 PS Claim 20; SEQ ID NO 165; 1503pp; English.  
 XX  
 CC The invention relates to 566 new isolated human polypeptides and their  
 CC encoding genes, sequences that are at least 95% identical to these or  
 CC sequences comprising one or more conservative substitutions in these. The  
 CC polypeptide, polynucleotide and antibodies against the polypeptides are  
 CC useful in diagnosing, treating or preventing NOVX-associated disorders,  
 CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,  
 CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,  
 CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.  
 CC The nucleic acids are further used as hybridization probes, in chromosome  
 CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The  
 CC polypeptides are also useful as vaccines. This sequence represents an  
 CC example of the nucleic acid sequence of the invention.  
 XX  
 SQ Sequence 6881 BP; 2297 A; 1362 C; 1301 G; 1921 T; 0 U; 0 Other;  
 Query Match 89.2%; Score 6340; DB 12; Length 6881;  
 Best Local Similarity 96.2%; Pred. No. 0;  
 Matches 6657; Conservative 0; Mismatches 215; Indels 48; Gaps 14;  
 QY 91 AAAAAATGTAATAAGATGGAATTTCTTATCATTTTCTTTTACTTTTATTTGGGACTTCA 150  
 DB 1 AAAAAATGTAATAAGATGGAATTTCTTATCATTTTCTTTTACTTTTATTTGGGACTTCA 60  
 QY 151 GAGACACAGGTGATGTTTCCAAATGTCGTCTCTGTTAGTACGATATACCATCTCT 210  
 DB 61 GAGACACAGGTAGATGTTTCCAAATGTCGTCTCTGTTAGTACGATATACCATCTCT 120  
 QY 211 TCAATTTCTACACATACACCTCCTCCTGTTTCTAGTAATAGTACACCAATGACAAA 270  
 DB 121 TCAATTTCTACACATACACCTCCTCCTGTTTCTAGTAATAGTACACCAATGACAAA 174  
 QY 271 CCAGGGCTCTCAGTCTTCTAGCCGGGAAAGAGTCCGATCTGCTGGGATCTTCTGTCT 330  
 DB 175 CCAGGGCTCTCAGTCTTCTAGCCGGGAAAGAGTCCGATCTGCTGGGATCTTCTGTCT 234  
 QY 331 TGAATACACACCTTAATCCAAATGGAAGATATATCTTCAATGTCAAATATAGGAA 390  
 DB 235 TGAATACACACCTTAATCCAAATGGAAGATATATCTTCAATGTCAAATATAGGAA 294  
 QY 391 GTTGTGCGTGGATGCAACAGTATATACACAGTCAGATCAAGCCGACAGCTCGNA 450  
 DB 295 GTTGTGCGTGGATGCAACAGTATATACACAGTCAGATCAAGCCGACAGCTCGNA 354

QY 451 GTTCTTCTTACTAATCTTAATCTCTCGAACACATATGAAATTAAGTTCTGCTGTAAC 510  
 DB 355 GTTCTTCTTACTAATCTTAATCTCTCGAACACATATGAAATTAAGTTCTGCTGTAAC 414  
 QY 511 AGTGTCTGGCATTTGGAGTGTGTTAGTATCCATTTCTTCCAAACTCAGAAAGTGTCTCA 570  
 DB 415 AGTGTCTGGCATTTGGAGTGTGTTAGTATCCATTTCTTCCAAACTCAGAAAGTGTCTCA 474  
 QY 571 GGAAAGTGGTGAATCTCAAGTTGAGGCTTCAACGGCTTTCAGCAGTTAAGCTGATTTGG 630  
 DB 475 GGAAAGTGGTGAATTTCAAGTGTGAGGCTGTC--CCGTTTCAGCAG-TAAGCTGATTTGG 531  
 QY 631 TATTTACTCTGGCAACCAATATGCGAAATTAACAGTTTCAAGATAGTGTCAAAATGCC 690  
 DB 532 TA--TACCTTCGCAACCAAAA-AAAAAAATTAACAGTTTCAAGATAGTGTCAAGCAATAAC 588  
 QY 691 AGAAGTGGGATAGTGTGAAAGATGTCTCAATCAGAGTAGAGGACATTTTGTGCTGGAAA 750  
 DB 589 AGAAGTGGGATAGTGTGAAAGATGTCTCAATCAGAGTAGAGTGTGCAATTTAAGTGTCTTC 648  
 QY 751 TTGCC--AGAAATGCAATGAGAAATAGTGAATCTTTTATGGAGTACAGCCAGCCCTTCT 807  
 DB 649 CTTCTCTTTCGACTGCAACGAGAAATAGTGAATCTTTTATGGAGTACAGCCAGCCCTTCT 708  
 QY 808 CCAACCTTTGGTAGAGTTACACCTTCATGCGGTACACACATTCATCAAGACAGTTGACA 867  
 DB 709 CCAACCTTTGGTAGAGTTACACCTTCATGCGGTACACACATTCATCAAGACAGTTGACA 768  
 QY 868 CAGAATGAGTACAGCTCTGTGGGAAAGAGCCTATCAGTTTGTAGTACACACTTGAGA 927  
 DB 769 CAGAATGAGTACAGCTCTGT--GAAAGAGCCTATCAGTTTGTAGTACACACTTGAGA 825  
 QY 928 CCTTATACAAACATATCTTTTGAAGTTTCAGCTGCTTACAACTGAAAGCAGTTATATTGAT 987  
 DB 826 CCTTATACAAACATATCTTTTGAAGTTTCAGCTGCTTACAACTGAAAGCAGTTATATTGAT 885  
 QY 988 AGTACGATTTGACAAACCAAGATTCAGTCTCTGAAGACCAACCAACCACTCGCTGTAACA 1047  
 DB 886 AGTACGATTTGACAAACCAAGATTCAGTCTCTGAAGACCAACCAACCACTCGCTGTAACA 945  
 QY 1048 GGCACATCAGAGGAAGTCTTTTCAATTTTATGGGACCCACCACTATAGTACAGGG 1107  
 DB 946 GGCACATCAGAGGAAGTCTTTTCAATTTTATGGGACCCACCACTATAGTACAGGG 1005  
 QY 1108 AAATTTAGTTATAGAGTTGAATTTATATGAGCCAT---CAGGTGCGATTTTGGATAACAGC 1164  
 DB 1006 AAATTTAGTTATAGAGTTGAATTTATATGAGCCATCAGCAGGTGCGATTTTGGATAACAGC 1065  
 QY 1165 ACAAAAGACCTCAAGTTTGCAATTCATTAACCTTAACCACTTAACCACTTAATGATGCTAT 1224  
 DB 1066 ACAAAAGACCTCAAGTTTGCAATTCATTAACCTTAACCACTTAACCACTTAATGATGCTAT 1125  
 QY 1225 ATTGGGCTGAAACAGTGCAGGGAGCTGGGCCCAAGTCAATATTTTCAGTATTTCACTCA 1284  
 DB 1126 ATTGGGCTGAAACAGTGCAGGGAGCTGGGCCCAAGTCAATATTTTCAGTATTTCACTCA 1185  
 QY 1285 CCAGATGTTTCAGGGGCGAGTGTTCATTTTCAACTTTGAGAGGTAGAATCCACGCAAGTA 1344  
 DB 1186 CCAGATGTTTCAGGGGCGAGTGTTCATTTTCAACTTTGAGAGGTAGAATCCACGCAAGTA 1245  
 QY 1345 AGAATTTCTGGAGAAACCAACGACCAACCAATGGAATTTTAAACCAATACCGAGTGAAA 1404  
 DB 1246 AGAATTTCTGGAGAAACCAACGACCAACCAATGGAATTTTAAACCAATACCGAGTGAAA 1305  
 QY 1405 GTGCTAGTTTCCAGAGACAGGAATATTTTGGAAATATCTTTGCTCCTCAGTGAATATAG 1464  
 DB 1306 GTGCTAGTTTCCAGAGACAGGAATATTTTGGAAATATCTTTGCTCCTCAGTGAATATAG 1365  
 QY 1465 TATATAAATGACCCCATGCTCCAGAAATTTGGAACATAGTAGAGCCCAATGTTAGATTA 1524  
 DB 1366 ---ATPAATGACCCCATGCTCCAGAAATTTGGAACATAGTAGAGCCCAATGTTAGATTA 1422  
 QY 1525 TATGAGGGTTTCAGCAGAGATGTCGTGACCTTCACTCACTGCTACATTTTATATATAAC 1584

|    |      |  |   |      |
|----|------|--|---|------|
| Db | 1423 | TATGAGGGTT   | CAGCAGAGATGTCGTCTGACCTTCACCTCACTTGCTCACTTATATATAAAC | 1482 |
| Qy | 1585 | AGCATCCAGATAAAACTTTCTCGAAGGAATAGAGCTGAAGCACCAGACTTCACCCAGTT      | 1644  |      |
| Db | 1483 | AGCATCCAGATAAAACTTTCTCGAAGGAATAGAGCTGAAGCACCAGACTTCACCCAGTT      | 1542  |      |
| Qy | 1645 | GTAACACAAAGGAATCAGTATATTACTGACATTCAGCTCAACAGCTGTCTTATGTTATC      | 1704  |      |
| Db | 1543 | GTAACACAGGAATCAGTATATTACTGACATTCAGCTGNAAGCTGACTTATGTTCTT         | 1602  |      |
| Qy | 1705 | AGGAGACTGTGTA CTTTTCACTGAGCA CATGATTAGTGTATCTGCTTTCA CACTCATGGGA | 1764  |      |
| Db | 1603 | ATCAGATTAAAGGAGATTTGGGGCTGAGACAATGGGGTTTTCTAGATATACAATCATG       | 1659  |      |
| Qy | 1765 | GAAGGACCAACACAGTCTTCAGTGTGTAGGACACGTCAGCAAGTGCACAGCTCCATTA       | 1824  |      |
| Db | 1660 | -----TCATCTCGAAGCAGGACAA TTTGACTTTCCCGAGCCCTTTGTGAGCCCA          | 1710  |      |
| Qy | 1825 | ATTATAAATCTATAAAATATTAGTTCCTTCATCTATTGTTTATATGGGATCCTCAGAA       | 1884  |      |
| Db | 1711 | AA TTT CAGATTACACATGTTACATAACAGAGTATTTTTTAC TGGGATCCTCCAGAT      | 1770  |      |
| Qy | 1885 | TATCCCAATGGAAAAATTAAC TCACTATA CGATTATATGCAATGGAATTTGGATACAACAGA | 1944  |      |
| Db | 1771 | CTGTATTTTTTTCATCATTAACCTTATCACTATTTTGGATGTTGAAACCAATCCAAGAT      | 1830  |      |
| Qy | 1945 | GCATTCAGATNACTACCATAGATAACAGCTTCTTCATAACAGGGTTAAAGAAATACACA      | 2004  |      |
| Db | 1831 | ATTATTTTAAGGACATTAACACAGTTTGCTCTCTGCTTATAGGGTTAAAGAAATACACA      | 1890  |      |
| Qy | 2005 | AAATACAAATGAGAGTGGCAGCTCAACCCACGATGGAGAAAGTTCCTTTGCTGTAAGAA      | 2064  |      |
| Db | 1891 | AAATACAAATGAGAGTGGCAGCTCAACCCACGTTGAGAAAGTTCCTTTGCTGTAAGAA       | 1950  |      |
| Qy | 2065 | AATGACATCTTTGTGAGAACTTCAGAAAGATGAACCGGAATCATCACCTCAAGATGTCGA     | 2124  |      |
| Db | 1951 | AATGACATCTTTGTGAGAACTTCAGAAAGATGAACCGGAATCATCACCTCAAGATGTCGA     | 2010  |      |
| Qy | 2125 | GTAATTCAGTTACCGCAGATGAATAAAGGTTGAAGTGGTCAACCCGGAAGCCCAAT         | 2184  |      |
| Db | 2011 | GTAATTCAGTTACCGCAGATGAATAAAGGTTGAAGTGGTCAACCCGGAAGCCCAAT         | 2070  |      |
| Qy | 2185 | GGGATCAATTTGCTTATGAAGTCTATATAAAATATAGATCTTTATATATGAAGAAC         | 2244  |      |
| Db | 2071 | GGGATCAATTTGCTTATGAAGTCTATATAAAATATAGATCTTTATATATGAAGAAC         | 2130  |      |
| Qy | 2245 | ACATCAACACAGACATAATATTAAGGAACCTTAAGACCTCACCCCTCTATAACATTTCT      | 2304  |      |
| Db | 2131 | ACATCAACACAGACATAATATTAAGGAACCTTAAGACCTCACCCCTCTATAACATTTCT      | 2190  |      |
| Qy | 2305 | GTAAGGTTTACACAGATTTGGTCACTGGCAATCAGGTATCTTCTTACTCTCTGTGAGG       | 2364  |      |
| Db | 2191 | GTAAGGTTTACACAGATTTGGTCACTGGCAATCAGGTATCTTCTTACTCTCTGTGAGG       | 2250  |      |
| Qy | 2365 | ACTTCGGAGACTGTGCTGATAGTGACACAGAAATATCACTTACAAAAATATTTCTTCT       | 2424  |      |
| Db | 2251 | ACTTCGGAGTCAGTGCTGATGATGACACAGAAATATCACTTACAAAAATATTTCTTCT       | 2310  |      |
| Qy | 2425 | GGAGAGATTGAGCTATCATTTCTTTCCCCCAAGTAGTCCCAATGGAATCATAAAAAATAT     | 2484  |      |
| Db | 2311 | GGAGAGATTGAGCTATCATTTCTTTCCCCCAAGTAGTCCCAATGGAATCATAAAAAATAT     | 2370  |      |
| Qy | 2485 | ACAATTTATCTCAAGAGAGTAAATGGAATATGAGGAAGAACTATAAATACAACCTCTTTA     | 2544  |      |
| Db | 2371 | ACAATTTATCTCAAGAGAGTAAATGGAATATGAGGAAGAACTATAAATACAACCTCTTTA     | 2430  |      |
| Qy | 2545 | ACCCAAACATTAAGTACTCAAGAAATATACCCAATATATCATTTGAGGTGCTGCTAGT       | 2604  |      |
| Db | 2431 | ACCCAAACATTAAGGTCTGAAGAAATATACCCAATATATCATTTGAGGTGCTGCTAGT       | 2490  |      |
| Qy | 2605 | ACACTGAAAGGTGAAGGAGTTTGGAGTGTCTCCCAATAAGTATACTGACGGAGGAAGATGCT   | 2664  |      |

|    |      |     |      |       |        |         |         |       |      |        |       |       |      |      |      |
|----|------|-----|------|-------|--------|---------|---------|-------|------|--------|-------|-------|------|------|------|
| Db | 2491 | ACA | CTCA | AAAGG | TGAAGG | AGTTCGG | AGTGTCC | CAAT  | AA   | GTATAT | CTG   | ACGG  | AGGA | GTAC | 2555 |
| Qy | 2665 | CTG | ATT  | CTCCC | CTCA   | AGACTT  | CTCTG   | TAAAA | CAGT | TGTCTG | TGTAC | CGGTG | CAAG | TTG  | 2724 |
| Db | 2551 | CTG | ATT  | CTCCC | CTCA   | AGACTT  | CTCTG   | TAAAA | CAGT | TGTCTG | TGTAC | CGGTG | CAAG | TTG  | 2610 |
| Qy | 2725 | TC  | TG   | CAAC  | CA     | CCCT    | TG      | GA    | CCAA | AT     | TAT   | CCT   | TTA  | TAC  | 2784 |
| Db | 2611 | TC  | TG   | CAAC  | CA     | CCCT    | TG      | GA    | CCAA | AT     | TAT   | CCT   | TTA  | TAC  | 2670 |
| Qy | 2785 | A   | A    | T     | A      | G       | A       | T     | C    | A      | T     | T     | A    | A    | 2844 |
| Db | 2671 | --- | A    | G     | A      | T       | C       | A     | T    | T      | A     | T     | A    | T    | 2727 |
| Qy | 2845 | G   | A    | T     | T      | A       | A       | T     | T    | G      | A     | A     | T    | T    | 2904 |
| Db | 2728 | G   | A    | T     | T      | A       | A       | T     | T    | G      | A     | A     | T    | T    | 2787 |
| Qy | 2905 | A   | C    | A     | G      | A       | G       | A     | C    | A      | A     | T     | T    | A    | 2964 |
| Db | 2788 | A   | C    | A     | G      | A       | G       | A     | C    | A      | A     | T     | T    | A    | 2844 |
| Qy | 2965 | G   | A    | T     | T      | A       | T       | A     | T    | A      | T     | A     | T    | A    | 3024 |
| Db | 2845 | G   | A    | T     | T      | A       | T       | A     | T    | A      | T     | A     | T    | A    | 2904 |
| Qy | 3025 | A   | A    | A     | C      | T       | A       | A     | T    | T      | A     | T     | A    | T    | 3084 |
| Db | 2905 | A   | A    | A     | C      | T       | A       | A     | T    | T      | A     | T     | A    | T    | 2964 |
| Qy | 3085 | A   | T    | C     | A      | G       | A       | A     | T    | T      | A     | T     | A    | T    | 3144 |
| Db | 2965 | A   | T    | C     | A      | G       | A       | A     | T    | T      | A     | T     | A    | T    | 3024 |
| Qy | 3145 | A   | T    | T     | A      | G       | A       | A     | T    | T      | A     | T     | A    | T    | 3204 |
| Db | 3025 | A   | T    | T     | A      | G       | A       | A     | T    | T      | A     | T     | A    | T    | 3084 |
| Qy | 3205 | G   | T    | T     | G      | A       | A       | T     | T    | A      | T     | A     | T    | A    | 3264 |
| Db | 3085 | G   | T    | T     | G      | A       | A       | T     | T    | A      | T     | A     | T    | A    | 3144 |
| Qy | 3265 | G   | A    | A     | G      | G       | T       | T     | G    | A      | A     | T     | T    | A    | 3324 |
| Db | 3145 | G   | A    | A     | G      | G       | T       | T     | G    | A      | A     | T     | T    | A    | 3204 |
| Qy | 3325 | T   | G    | G     | T      | C       | C       | A     | C    | G      | G     | T     | C    | A    | 3384 |
| Db | 3205 | T   | G    | G     | T      | C       | C       | A     | C    | G      | G     | T     | C    | A    | 3264 |
| Qy | 3385 | C   | A    | G     | A      | C       | T       | T     | G    | A      | C     | C     | A    | C    | 3444 |
| Db | 3265 | C   | A    | G     | A      | C       | T       | T     | G    | A      | C     | C     | A    | C    | 3324 |
| Qy | 3445 | A   | A    | T     | C      | T       | G       | A     | A    | A      | A     | T     | T    | A    | 3504 |
| Db | 3325 | A   | A    | T     | C      | T       | G       | A     | A    | A      | A     | T     | T    | A    | 3384 |
| Qy | 3505 | T   | C    | T     | G      | A       | T       | A     | T    | A      | T     | A     | T    | A    | 3564 |
| Db | 3385 | T   | C    | T     | G      | A       | T       | A     | T    | A      | T     | A     | T    | A    | 3444 |
| Qy | 3565 | C   | C    | A     | A      | T       | A       | T     | A    | T      | A     | T     | A    | T    | 3624 |
| Db | 3445 | C   | C    | A     | A      | T       | A       | T     | A    | T      | A     | T     | A    | T    | 3504 |
| Qy | 3625 | C   | C    | A     | G      | T       | A       | A     | G    | T      | T     | A     | T    | A    | 3684 |
| Db | 3505 | C   | C    | A     | G      | T       | A       | A     | G    | T      | T     | A     | T    | A    | 3564 |
| Qy | 3685 | A   | A    | T     | T      | A       | T       | T     | A    | T      | T     | A     | T    | A    | 3744 |
| Db | 3565 | A   | A    | T     | T      | A       | T       | T     | A    | T      | T     | A     | T    | A    | 3624 |





5785 CAGAAAGAGGTGGCACAATCTCTCTCAGGATGCGAGAAATATTGACACTAAATTTGAAG 5844  
5965 CTGGATCAGCTCATCAGTGGCAGACCTGGAACTGGAAGCAGAGGATTAACGGCGCCA 6024  
5845 CTGGATCAGCTCATCAGTGGCAGACCTGGAACTGGAAGCAGAGGATTAACGGCGCCA 5904  
6025 ATAAGCAAGAAATCCTTCTGCAACATGTTGAAGAGCTTTGACAAACCAACCACTTAAG 6084  
5905 ATAAGCAAGAAATCCTTCTGCAACATGTTGAAGAGCTTTGACAAACCAACCACTTAAG 5964  
6085 TTTCAAGAGAAATTTTCGGAATTTACCAAAATTTCTTCAGGATCTTTCTCAACTGATGCT 6144  
5965 TTTCAAGAGAAATTTTCGGAATTTACCAAAATTTCTTCAGGATCTTTCTCAACTGATGCT 6024  
6145 GATCTGCTTGGAAATAGAGCAAAACCGTTTCCCAACATATAAACCATAATATAAT 6204  
6025 GATCTGCTTGGAAATAGAGCAAAACCGTTTCCCAACATATAAACCATAATATAAT 6081  
6205 AACAGAGTAAGCTGATAGCTGACGCTAGTGTTCAGGTTCCGATTTATTAATGCGAGC 6264  
6082 AACAGAGTAAGCTGATAGCTGACGCTAGTGTTCAGGTTCCGATTTATTAATGCGAGC 6141  
6265 TATATTTCTGGTTTATTTATGTCCTCAATGAATTTATTTGCTCAAGTCCACTACCAGGA 6324  
6142 TATATTTCTGGTTTATTTATGTCCTCAATGAATTTATTTGCTCAAGTCCACTACCAGGA 6201  
6325 ACAGTTGGAGATTTTGGAGATGTTGGTGGGAAACAGGGGCAAAACATTAATGCTA 6384  
6202 ACAGTTGGAGATTTTGGAGATGTTGGTGGGAAACAGGGGCAAAACATTAATGCTA 6261  
6385 ACACAGTGTTTGAAAAGAGCGATCAGATGCTATGCTATTTGGCAGAGCAACACAG 6444  
6262 ACACAGTGTTTGAAAAGAGCGATCAGATGCTATGCTATTTGGCAGAGCAACACAG 6321  
6445 CCAGTTTACTGCTCTTGGAGATAGTGAATTAACAAGCTTAATGGAGGATTTCAATAGAT 6504  
6322 CCAGTTTACTGCTCTTGGAGATAGTGAATTAACAAGCTTAATGGAGGATTTCAATAGAT 6381  
6505 TGGAATCAGGATCTGAAAATTTGAAAGCATGCGGATGATGATCTGTTGACAGTGT 6564  
6382 TGGAATCAGGATCTGAAAATTTGAAAGCATGCGGATGATGATCTGTTGACAGTGT 6441  
6565 AACTTTTACTGCTGCGCAGAGATGCGGTTCTTGAGAACAGCGCCCTCTAATTCACTTT 6624  
6442 AACTTTTACTGCTGCGCAGAGATGCGGTTCTTGAGAACAGCGCCCTCTAATTCACTTT 6501  
6625 GTGAAGTTGTTGAGAGCAGGCGACATGACACCACTATGATTTGTTCACTGCACT 6684  
6502 GTGAAGTTGTTGAGAGCAGGCGACATGACACCACTATGATTTGTTCACTGCACT 6561  
6685 GTGGAGTTGGAAGAACTGGAGTTTATTTGCTTGGACCAATTTAAACAATATAAT 6744  
6562 GTGGAGTTGGAAGAACTGGAGTTTATTTGCTTGGACCAATTTAAACAATATAAT 6621  
6745 GACCATGATTTGTCATATATATGACATGATGCTGACTGAGTGAAGTGAAGATGTC 6804  
6622 GACCATGATTTGTCATATATATGACATGATGCTGACTGAGTGAAGTGAAGATGTC 6681  
6805 ATGGTGCAGAACTGCGCAGATATATCTTTTACACCACTGATCTTGGATCTTTATCA 6864  
6682 ATGGTGCAGAACTGCGCAGATATATCTTTTACACCACTGATCTTGGATCTTTATCA 6741  
6865 AATAAGGAGTAATCAGGCCATCTGTTTGTGTTTAACTATTCAGCACTCAGAGATGAC 6924  
6742 AATAAGGAGTAATCAGGCCATCTGTTTGTGTTTAACTATTCAGCACTCAGAGATGAC 6801  
6925 TCTTTGAGCCCATGAA---GCTGATGTTGAGCTTGAATGGAAGAAACCACTATGTA 6981  
6802 TCTTTGAGCCCATGGAAGTGTGTGTTGAGCTTGAATGGAAGAAACCACTATGTA 6861  
6982 ATATTCCAGCAAGGATAC 7001

Db 6862 ATATTCCAGACCAAGGATAC 6881  
RESULT 10  
AAD56413  
ID AAD56413 standard; DNA; 5877 BP.  
XX  
AC AAD56413;  
XX  
DT 07-AUG-2003 (first entry)  
XX  
DE Human receptor tyrosine kinase DNA #3.  
XX  
KW Human; receptor tyrosine phosphatase; diabetes; obesity; CNS disorder;  
KW multiple sclerosis; epilepsy; chronic obstructive pulmonary disease;  
KW COPD; Parkinson's disease; stroke; cardiovascular disorder; arrhythmia;  
KW congestive heart failure; myocardial infarction; ischaemic heart disease;  
KW gene therapy; anorectic; cardiant; neuroprotective; anticonvulsant;  
KW cerebroprotective; vasotropic; antiarrhythmic; receptor; enzyme; gene;  
KW ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
CDS 1..5877  
FT /\*tag= a  
FT /product= "Human receptor tyrosine phosphatase protein"  
FT /note= "No start and stop codon"  
FT /partial  
XX  
PN WO2003033688-A1.  
XX  
PD 24-APR-2003.  
XX  
PF 14-OCT-2002; 2002WO-EP011473.  
XX  
PR 16-OCT-2001; 2001US-0329329P.  
XX  
PA (FARB ) BAYER AG.  
XX  
PI Liou J;  
XX  
DR WPI; 2003-403215/38.  
DR P-PSDB; AAE37322.  
XX  
PT Novel polynucleotides encoding human receptor tyrosine phosphatase  
PT polypeptides, useful for treating diabetes, CNS disorders, obesity,  
PT chronic obstructive pulmonary disease and cardiovascular disorders.  
XX  
PS Disclosure; Fig 7; 163pp; English.  
XX  
CC The present invention relates to receptor tyrosine phosphatase proteins  
CC and polynucleotides encoding them. Sequences of the invention are useful  
CC in the preparation of medicaments for modulating the activity of receptor  
CC tyrosine phosphatase in disease such as diabetes, obesity, CNS disorders  
CC (multiple sclerosis, epilepsy, Parkinson's disease and stroke), chronic  
CC obstructive pulmonary disease (COPD) and cardiovascular disorders (e.g.  
CC congestive heart failure, myocardial infarction, ischaemic heart disease  
CC and arrhythmia). They are also used in gene therapy. the present sequence  
CC is human receptor tyrosine kinase DNA  
XX  
SQ Sequence 5877 BP; 2030 A; 1157 C; 1087 G; 1603 T; 0 U; 0 Other;  
Query Match 79.1%; Score 5619.4; DB 9; Length 5877;  
Best Local Similarity 97.6%; Pred. No. 0;  
Matches 5736; Conservative 0; Mismatches 111; Indels 30; Gaps 2;  
QY 1129 TTATATGACCACTCAGTGGCATTGTTGGATACAGCACAACCAAGCTCAAGTTTGCAATC 1188  
Db 4 TTAGAGGAGAACTAACAAACAGAGGACATCCACACCAAAACCCATCTGTACATCAC 63  
QY 1189 ACTAACCTAACCACTATTACAAATGATGATGTTATTTATTTGCGGCTGAAACCACTGCGAGG 1248

Db 64 CATCATCAAGACCAAAAGTAGATAAAACCAAAAGTGGGAAAAAACAAGACGACGAAA 123  
Qy 1249 ACTGGGCCCCAAGTCAAAATATTTTCAGTATTCATCTCCACCGAGATGTTTCAAGGGGCAAGTGT 1308  
Db 124 ACTGGAACTCTAAAAAGACAGACACCTCTCTCTCTCCAAAGGATGGCAGGGCAGTGT 183  
Qy 1309 GATTTCACCTTCGACAGGTAGAAATCCAGCAAGTAAAGTAAATTAATTTGGAGAAACACCA 1368  
Db 184 GATTTCACCTTCGACAGGTAGAAATCCAGCAAGTAAAGTAAATTAATTTGGAGAAACACCA 243  
Qy 1369 CAACCAAAATGGAATTTATTAACCAATACCGAGTGAAGTGTCTTCCAGAGACAGGAATA 1428  
Db 244 CAACCAAAATGGAATTTATTAACCAATACCGAGTGAAGTGTCTTCCAGAGACAGGAATA 303  
Qy 1429 ATTTGGAAATTAATTTGCTCACTGGAATAATAGATATATAAATGACCCATCGGTCCA 1488  
Db 304 ATTTGGAAATTAATTTGCTCACTGGAATAATAGATATATAAATGACCCATCGGTCCA 363  
Qy 1489 GAAATTTGNAACATAGTAGGCCAATGTTAGGATTTATGAGGGTTTCAGCAGAGATGTCG 1548  
Db 364 GAAATTTGNAACATAGTAGGCCAATGTTAGGATTTATGAGGGTTTCAGCAGAGATGTCG 423  
Qy 1549 TCTGACCTTCACCTTCGCTACATTTATATAAACAAGCCATCCAGATATAAACTTTCT 1608  
Db 424 TCTGACCTTCACCTTCGCTACATTTATATAAACAAGCCATCCAGATATAAACTTTCT 483  
Qy 1609 GCAAGGAATAGAGCTGAAGACCAAGCTTCAACAGTTGTAACTAACAAGGAATCAGTATAT 1668  
Db 484 GCAAGGAATAGAGCTGAAGACCAAGCTTCAACAGTTGTAACTAACAAGGAATCAGTATAT 543  
Qy 1669 ACTGACATTCGACGTGAACAGCTGTCTTATGTTTATGAGGACTTGTACCTTTCAGTGAG 1728  
Db 544 ACTGACATTCGACGTGAACAGCTGTCTTATGTTTATGAGGACTTGTACCTTTCAGTGAG 603  
Qy 1729 CACATGATTAGTGTATCTGCTTTCACCATCATCGGAGAGGACCAACAAGTTCTCAGT 1788  
Db 604 CACATGATTAGTGTATCTGCTTTCACCATCATCGGAGAGGACCAACAAGTTCTCAGT 663  
Qy 1789 GTTAGGACAGCTCAGCAAGTGCACAGCTCCATTTAAATATATAAAATATTAAGT 1848  
Db 664 GTTAGGACAGCTCAGCAAGTGCACAGCTCCATTTAAATATATAAAATATTAAGT 723  
Qy 1849 TCTTCATCTATTTGTTATTTGGGATCTTCAGATATATCCCAATCGGAAAAATTAACCTAC 1908  
Db 724 TCTTCATCTATTTGTTATTTGGGATCTTCAGATATATCCCAATCGGAAAAATTAACCTAC 783  
Qy 1909 TATACGATTTATGCAATGGAATTTGGATACAAACAGAGCATTCAGATAACTACATAGAT 1968  
Db 784 TATACGATTTATGCAATGGAATTTGGATACAAACAGAGCATTCAGATAACTACATAGAT 843  
Qy 1969 AACAGCTTTTCTCAACAGGTTTAAAGAAATACAAAATACAAAATGAGGTGGCAGCC 2028  
Db 844 AACAGCTTTTCTCAACAGGTTTAAAGAAATACAAAATACAAAATGAGGTGGCAGCC 903  
Qy 2029 TCAACCCAGATGGAGAAAGTTCTTTGCTCGAAGAAATGAACATCTTTGTGAGAACTTCA 2088  
Db 904 TCAACCCAGATGGAGAAAGTTCTTTGCTCGAAGAAATGAACATCTTTGTGAGAACTTCA 963  
Qy 2089 GAAGATGAAACCGGAATCATCACCCTCAAGATGTCGAAGTAAATGATGTTACCCGAGATGAA 2148  
Db 964 GAAGATGAAACCGGAATCATCACCCTCAAGATGTCGAAGTAAATGATGTTACCCGAGATGAA 1023  
Qy 2149 ATAAGGTTGAAGTGGTCACCAACCCGAAAGCCCAATGGGATCAATTTGCTTATGAAGTG 2208  
Db 1024 ATAAGGTTGAAGTGGTCACCAACCCGAAAGCCCAATGGGATCAATTTGCTTATGAAGTG 1083  
Qy 2209 CTATATAAAATATAGACTTTTATATATGAAGAACACATCAACACAGACATATAATTA 2268  
Db 1084 CTATATAAAATATAGACTTTTATATATGAAGAACACATCAACACAGACATATAATTA 1143  
Qy 2269 AGGAACCTTAAGACCTCACACCCCTCTATAACATTTCTGTAAGGTCTTACACGAGTTGGT 2328  
Db 1144 AGGAACCTTAAGACCTCACACCCCTCTATAACATTTCTGTAAGGTCTTACACGAGTTGGT 1203

Qy 2329 CATGGCAATCAGGTATCTTTCTTTTACTCTCTGTAAAGACTTTCGGAGACTGTGCTGTATAGT 2388  
Db 1204 CATGGCAATCAGGTATCTTTCTTTTACTCTCTGTAAAGACTTTCGGAGACTGTGCTGTATAGT 1263  
Qy 2389 GCACCAAGAAATATACATTTACAAAATATTTCTTCTGGAGAGATTCAGCTATCATTTCCCT 2448  
Db 1264 GCACCAAGAAATATACATTTACAAAATATTTCTTCTGGAGAGATTCAGCTATCATTTCCCT 1323  
Qy 2449 CCCCCAAGTAGTCCCAATGGAATCATAAAAAATATACAATTTATCTCAAGAGAACTAAT 2508  
Db 1324 CCCCCAAGTAGTCCCAATGGAATCATAAAAAATATACAATTTATCTCAAGAGAACTAAT 1383  
Qy 2509 GGAATAGGAAAGAACTATATAATACAACCTCTTTTAAACCAAAACATTTAAAGTACTGAAG 2568  
Db 1384 GGAATAGGAAAGAACTATATAATACAACCTCTTTTAAACCAAAACATTTAAAGTACTGAAG 1443  
Qy 2569 AAATATACCCCAATATATCATTCAGGTGCTCTGTAGTACACTGAAAGGTGAAGAGTTCCG 2628  
Db 1444 AAATATACCCCAATATATCATTCAGGTGCTCTGTAGTACACTGAAAGGTGAAGAGTTCCG 1503  
Qy 2629 AGTGTCTCCCATAGTATACCTGACGGAAGAGTGTCTCTGTATCTCCCTCAAGACTTC 2688  
Db 1504 AGTGTCTCCCATAGTATACCTGACGGAAGAGTGTCTCTGTATCTCCCTCAAGACTTC 1563  
Qy 2689 TCTGTAAAAACAGTTGTCTGCTGTCACGGTGAAGTGTCTATGGCAACCAACCCCTGGAGCCA 2748  
Db 1564 TCTGTAAAAACAGTTGTCTGCTGTCACGGTGAAGTGTCTATGGCAACCAACCCCTGGAGCCA 1623  
Qy 2749 AATGGAATTTATCTTTTATACACAGTTTATGCTGGAATAGATCATTTAAAAACTAT 2808  
Db 1624 AATGGAATTTATCTTTTATACACAGTTTATGCTGGAATAGATCATTTAAAAACTAT 1683  
Qy 2809 AATGTCACCTGAAACATCATTTGGAGTTATCAGATTTTGGATTTAAATGTTGAATACAGTGT 2868  
Db 1684 AATGTCACCTGAAACATCATTTGGAGTTATCAGATTTTGGATTTAAATGTTGAATACAGTGT 1743  
Qy 2869 TATGTAAACAGCTAGACCAAGATTTGGTGTATGGGAAAAACAGGAAGCAATATCATTTAGT 2928  
Db 1744 TATGTAAACAGCTAGACCAAGATTTGGTGTATGGGAAAAACAGGAAGCAATATCATTTAGT 1803  
Qy 2929 CAAACACAGAGGGGAGCACCAAGCGATCTCCCAAGATGTTTATTTATCCAAACCTCAGT 2988  
Db 1804 CAAACACAGAGGGGAGCACCAAGCGATCTCCCAAGATGTTTATTTATGAAACCTCAGT 1863  
Qy 2989 TCTTCATCAATTAATTTCTTTCTGGACACCTCTCTCAAAACCTAAATGGGATTAACAATAT 3048  
Db 1864 TCTTCATCAATTAATTTCTTTCTGGACACCTCTCTCAAAACCTAAATGGGATTAACAATAT 1923  
Qy 3049 TACTCTGTTTATTAACAGAAATACCTTCAGGTACTTTTATGCGAGAAATTTTACCTCCATGAA 3108  
Db 1924 TACTCTGTTTATTAACAGAAATACCTTCAGGTACTTTTATGCGAGAAATTTTACCTCCATGAA 1983  
Qy 3109 CTAACCAATGACTTTGACAAATATGCTGTATCCAAATTTAGTAACTGACAAATTTTC 3168  
Db 1984 GTAACCAATGACTTTGACAAATATGCTGTATCCAAATTTAGTAACTGACAAATTTTC 2043  
Qy 3169 AGCTACTATACATTTTGGTTTAAACAGCAAGTACTCTCAGTTGGAAATGGGAAATAAAGCAGT 3228  
Db 2044 AGCTACTATACATTTTGGTTTAAACAGCAAGTACTCTCAGTTGGAAATGGGAAATAAAGCAGT 2103  
Qy 3229 GACATCATTTGAAAGTATACACAGATCAAGACATACCTTGAAAGGGTGTGTTGGAAACCTGACT 3288  
Db 2104 GACATCATTTGAAAGTATACACAGATCAAGACATACCTTGAAAGGGTGTGTTGGAAACCTGACT 2163  
Qy 3289 TAGGAATCCATTTGCTCAACTGCATTAATGTAAGCTGGGTCCACCGGCTCAACCAAC 3348  
Db 2164 TAGGAATCCATTTGCTCAACTGCATTAATGTAAGCTGGGTCCACCGGCTCAACCAAC 2223  
Qy 3349 GGTCTAGTCTCTCTACTATGTTTCACTGATCTTTACAGCAGACTCTCTGCGCATGTGAGACCA 3408  
Db 2224 GGTCTAGTCTCTCTACTATGTTTCACTGATCTTTACAGCAGACTCTCTGCGCATGTGAGACCA 2283

|    |      |  |      |
|----|------|--|------|
| Qy | 3409 | CCTCTTGTTACATATGAGAGAGCATATATATTTTGATTAATCTCGGAAAAATACACTGATTTAT | 3468 |
| Db | 2284 | CCTCTTGTTACATATGAGAGAGCATATATTTTGATTAATCTCGGAAAAATACACTGATTTAT   | 2343 |
| Qy | 3469 | ATATTAATAATTACTCCATCAACAGAAAAGGGATTCTCTGATACCTATATCTGCCAGCTA     | 3528 |
| Db | 2344 | ATATTAATAATTACTCCATCAACAGAAAAGGGATTCTCTGATACCTATATCTGCCAGCTA     | 2403 |
| Qy | 3529 | TACATCAAGACTGAAGAAGATGTCCCGAAGAACTTCACCAATATATCAACACTTTTAAAAAC   | 3588 |
| Db | 2404 | TACATCAAGACTGAAGAAGATGTCCCGAAGAACTTCACCAATATATCAACACTTTTAAAAAC   | 2463 |
| Qy | 3589 | CTTTCTCTACTCTCAGTTCCTTTATCATGGGATCCCGCAGTAAGCCAAATGTGTGCAATA     | 3648 |
| Db | 2464 | CTTTCTCTACTCTCAGTTCCTTTATCATGGGATCCCGCAGTAAGCCAAATGTGTGCAATA     | 2523 |
| Qy | 3649 | ATAAGTTATGATTTTAACTTTTACAAGGACCAATGAAAATTTATTCCTTCATTACTTCTGAT   | 3708 |
| Db | 2524 | ATAAGTTATGATTTTAACTTTTACAAGGACCAATGAAAATTTATTCCTTCATTACTTCTGAT   | 2583 |
| Qy | 3709 | AAATACATATATTTGGAAGAGCTTCACCACTTTTACATTTATATAGCTTTTTTGTGCGCGCA   | 3768 |
| Db | 2584 | AAATACATATATTTGGAAGAGCTTCACCACTTTTACATTTATATAGCTTTTTTGTGCGCGCA   | 2643 |
| Qy | 3769 | AGAACTAGAAAAGGACTTGTGCTTCAGTATCTTTTTTCTTTTACACAGATGATGAGTG       | 3828 |
| Db | 2644 | AGAACTAGAAAAGGACTTGTGCTTCAGTATCTTTTTTCTTTTACACAGATGATGAGTG       | 2703 |
| Qy | 3829 | CCGTTAGCACCTCCACAAAAATTTGACATTTTAATCAACTGTACTTCAGACTTTGTATGGCTG  | 3888 |
| Db | 2704 | CCGTTAGCACCTCCACAAAAATTTGACATTTTAATCAACTGTACTTCAGACTTTGTATGGCTG  | 2763 |
| Qy | 3889 | AAATGGAGCCCAAGCTCCTCTCCAGGTGGTATTTGTTAAAGATATATAGTTTTTAAAAATTCAT | 3948 |
| Db | 2764 | AAATGGAGCCCAAGCTCCTCTCCAGGTGGTATTTGTTAAAGATATATAGTTTTTAAAAATTCAT | 2823 |
| Qy | 3949 | GAACTAGAAACTGACACTATATATTAAGAAATATATCAGGATTTTAAAACTGAAGCCAAA     | 4008 |
| Db | 2824 | GAACTAGAAACTGACACTATATATTAAGAAATATATCAGGATTTTAAAACTGAAGCCAAA     | 2883 |
| Qy | 4009 | CTTGTTGGACTGGAAACAGTGCAGACACTACTATATCCGTGTATCTGGTTCAACAAGTT      | 4068 |
| Db | 2884 | CTTGTTGGACTGGAAACAGTGCAGACACTACTATATCCGTGTATCTGGTTCAACAAGTT      | 2943 |
| Qy | 4069 | GGAATGGCAATCAATTTAGTAAATGTAGTAAATTTCAACCCAAAGAAATCAGTTCAGAT      | 4128 |
| Db | 2944 | GGAATGGCAATCAATTTAGTAAATGTAGTAAATTTCAACCCAAAGAAATCAGTTCAGAT      | 3003 |
| Qy | 4129 | GTCTGTGCAAGATATGCAGTGCATGGCAACTAGCTGGCAGTCAGTTTTAGTGAATGGAT      | 4188 |
| Db | 3004 | GTCTGTGCAAGATATGCAGTGCATGGCAACTAGCTGGCAGTCAGTTTTAGTGAATGGAT      | 3063 |
| Qy | 4189 | CCACCCAAAAAGGCAATGGAATTAATAACGAGTATATGGTAAACAGTTGAAAAGGAATTCCT   | 4248 |
| Db | 3064 | CCACCCAAAAAGGCAATGGAATTAATAACGAGTATATGGTAAACAGTTGAAAAGGAATTCCT   | 3123 |
| Qy | 4249 | ACAAAAGTTTCTCCCGAAGTCACTGTACATTTTCAATAAGCTTCTTGCCCAATACCTCA      | 4308 |
| Db | 3124 | ACAAAAGTTTCTCCCGAAGTCACTGTACATTTTCAATAAGCTTCTTGCCCAATACCTCA      | 3183 |
| Qy | 4309 | TATGCTTTTAAAGTAAGAGCTTCAACCTCAGCTGGTGAAGGTGATGAAAGCAATGCCAT      | 4368 |
| Db | 3184 | TATGCTTTTAAAGTAAGAGCTTCAACCTCAGCTGGTGAAGGTGATGAAAGCAATGCCAT      | 3243 |
| Qy | 4369 | GTGAGCACACTPACTGAAAACAGTTCCAGTGTTCGCCAATAATATGCTTTTTCTGATGTT     | 4428 |
| Db | 3244 | GTGAGCACACTPACTGAAAACAGTTCCAGTGTTCGCCAATAATATGCTTTTTCTGATGTT     | 3303 |
| Qy | 4429 | CAGTCAACTAGTGCAACATTTGACATGATGAAAGCTGACACTATCTCTTGCTACTTTCAA     | 4488 |
| Db | 3304 | CAGTCAACTAGTGCAACATTTGACATGATGAAAGCTGACACTATCTCTTGCTACTTTCAA     | 3363 |
| Qy | 4489 | AAATACAAAAATTAACACTCAACTTCGTGCTCAAAAAATGCAAGAAATGGGAATCCGAAGAA   | 4548 |

Db 4444 GGCTTTCCTAACCCCTCCATGTACAGAGGAAACAAAGTTTATAGTGCATGAGAAATC 4503  
Qy 5629 TACATCATAGTGTCTGATTAATGATGATGATTCCTGCAATGAGACAAATTTTCAAT 5688  
Db 4504 TACATCATAGTGTCTGATTAATGATGATTCCTGCAATGAGACAAATTTTCAAT 4563  
Qy 5689 GGACCACTGAAACCAAAAGCAATCTTATTTAAATTTAGCTCAAAATATTGGGA 5748  
Db 4564 GGACCACTGAAACCAAAAGCAATCTTATTTAAATTTAGCTCAAAATATTGGGA 4623  
Qy 5749 CAATTTACTGACTCTGATTAATCTGACCCCTGTTAAGACTTTAGGGAAGGACTTTTCAGAA 5808  
Db 4624 CAATTTACTGACTCTGATTAATCTGACCCCTGTTAAGACTTTAGGGAAGGACTTTTCAGAA 4683  
Qy 5809 AGAACCGTAGAGATCATCTTTCCGTCACATTTGCTGATCTTTCAATAATTTCTCTTGA 5868  
Db 4684 AGAACCGTAGAGATCATCTTTCCGTCACATTTGCTGATCTTTCAATAATTTCTCTTGA 4743  
Qy 5869 ACAGCTATTCTTTCATTTTCGAAGATTCGACAGAGCAAGAAAGGTCGCATCTCT 5928  
Db 4744 ACAGCTATTCTTTCATTTTCGAAGATTCGACAGAGCAAGAAAGGTCGCATCTCT 4803  
Qy 5929 CCTCAGGATCGAGAAATTTATGACACTAAATTCGAAGCTGGATCAGCTCATCAGTGGCA 5988  
Db 4804 CCTCAGGATCGAGAAATTTATGACACTAAATTCGAAGCTGGATCAGCTCATCAGTGGCA 4863  
Qy 5989 GACCTGGAACTGAAGGACGAGATTAACGCG-----G 6021  
Db 4864 GACCTGGAACTGAAGGACGAGATTAACGCGTTACTTAGTTATAGAAATCCATCAAG 4923  
Qy 6022 CCAATAGCAAGAAATCTCTCTGCAACATGTTGGAAGAGCTTTGCAACAAACCACTTA 6081  
Db 4924 CCAATAGCAAGAAATCTCTCTGCAACATGTTGGAAGAGCTTTGCAACAAACCACTTA 4983  
Qy 6082 AGTTTTCAGAGAAATTTTCGGAATTTACCAAAATTTCTTCAGGATCTTTCTCACTGAT 6141  
Db 4984 AGTTTTCAGAGAAATTTTCGGAATTTACCAAAATTTCTTCAGGATCTTTCTCACTGAT 5043  
Qy 6142 GCTGATCTGCTTGGAAATAGAGCAAAACCGTTTCCCAACATAAAACCATATAAT 6201  
Db 5044 GCTGATCTGCTTGGAAATAGAGCAAAACCGTTTCCCAACATAAAACCAT---ATAAT 5100  
Qy 6202 AATAACAGATTAAGCTGATAGCTGACGGTAGTGTTCAGGTTCCGATTAATTAATGCC 6261  
Db 5101 AATAACAGATTAAGCTGATAGCTGACGGTAGTGTTCAGGTTCCGATTAATTAATGCC 5160  
Qy 6262 AGCTATATTCTGTTTATTTATGTCCTCAAAATGATTTATGCTACTCAAGTCCACTACCA 6321  
Db 5161 AGCTATATTCTGTTTATTTATGTCCTCAAAATGATTTATGCTACTCAAGTCCACTACCA 5220  
Qy 6322 GGAAACAGTTGGAGATTTTGGAGAAATGTTGGGAAACAGGGCAAAACATTTAGTAATG 6381  
Db 5221 GGAAACAGTTGGAGATTTTGGAGAAATGTTGGGAAACAGGGCAAAACATTTAGTAATG 5280  
Qy 6382 CTAACACAGTGTTTTGAAGAAAGGACGGATCAGATGCCATCAGTATTTGGCAGAGGCAAC 6441  
Db 5281 CTAACACAGTGTTTTGAAGAAAGGACGGATCAGATGCCATCAGTATTTGGCAGAGGCAAC 5340  
Qy 6442 AAGCCAGTTACTGTCTTTGGAGATATAGTATTAACAAAGTATGAGGATGTTCAATA 6501  
Db 5341 AAGCCAGTTACTGTCTTTGGAGATATAGTATTAACAAAGTATGAGGATGTTCAATA 5400  
Qy 6502 GATTGACTATCAGGATCTGAAATTTGAAAGCATGGGATTTGCAATGATGTTGCACAG 6561  
Db 5401 GATTGACTATCAGGATCTGAAATTTGAAAGCATGGGATTTGCAATGATGTTGCACAG 5460  
Qy 6562 TGTAATCTTTACTGCTGCGCAGAGCATGGGGTTCTTGAGAAACAGCGCCCTCTTAATTCAC 6621  
Db 5461 TGTAATCTTTACTGCTGCGCAGAGCATGGGGTTCTTGAGAAACAGCGCCCTCTTAATTCAC 5520  
Qy 6622 TTTGTGAAGTTGTTTCAGCAAGCAGGGCAGATGACCAACCTATGATTTGTTCACTGC 6681  
Db 5521 TTTGTGAAGTTGTTTCAGCAAGCAGGGCAGATGACCAACCTATGATTTGTTCACTGC 5580

Qy 6682 AGTGTCTGGAGTTGGAGAACTGGAGTTTATTTATTTCTCTGGACCATTTTAAACAAATATA 6741  
Db 5581 AGTGTCTGGAGTTGGAGAACTGGAGTTTATTTATTTCTCTGGACCATTTTAAACAAATATA 5640  
Qy 6742 AATGACCATGATTTTGTGGATATATATGACTAGTAGCTGAACTGAGAAGTGAAGAAATG 6801  
Db 5641 AATGACCATGATTTTGTGGATATATATGACTAGTAGCTGAACTGAGAAGTGAAGAAATG 5700  
Qy 6802 TGCATGTTGTCAGAACTGCGCAGATATATCTTTTATACCAAGTGCATTTCTGGATCTCTTA 6861  
Db 5701 TGCATGTTGTCAGAACTGCGCAGATATATCTTTTATACCAAGTGCATTTCTGGATCTCTTA 5760  
Qy 6862 TCAATATAGGAAAGTAAATCAGCCCATCTGTTTGTAACTATTCAGCATTTCAGAAATG 6921  
Db 5761 TCAATATAGGAAAGTAAATCAGCCCATCTGTTTGTAACTATTCAGCATTTCAGAAATG 5820  
Qy 6922 GACTCTTTGACGCCCATGGAAGTGTGATGTTGAGCTTGAATGGAAGAAACCACTATG 6978  
Db 5821 GACTCTTTGACGCCCATGGAAGTGTGATGTTGAGCTTGAATGGAAGAAACCACTATG 5877

## RESULT 11

ADH41620

ID ADH41620 standard; DNA; 2739 BP.

XX AC

XX AC

XX 25-MAR-2004 (firet entry)

XX XX

XX Novel human nucleic acid NOV15e.

ds; gene; cardiovascular; antiarteriosclerotic; hypotensive; cytostatic;  
anorectic; antidiabetic; immunosuppressive; anti-HIV; neuroprotective;  
neurotic; antiparkinsonian; antisthmatic; antiinfertility;  
cardiomyopathy; atherosclerosis; hypertension; cancer; obesity; diabetes;  
AIDS; multiple sclerosis; graft-versus-host disease; Alzheimer's disease;  
Parkinson's disease; asthma; fertility disorder; chromosome mapping;  
tissue typing; preventive medicine; pharmacogenomic; vaccine.

XX Homo sapiens.

XX WO2003102159-A2.

XX 11-DEC-2003.

XX 04-JUN-2003; 2003WO-US017573.

XX 04-JUN-2003; 2002US-0385490P.

XX 04-JUN-2003; 2002US-0385615P.

XX 04-JUN-2003; 2002US-0385755P.

XX 05-JUN-2003; 2002US-0386041P.

XX 06-JUN-2003; 2002US-0386355P.

XX 06-JUN-2003; 2002US-0386357P.

XX 06-JUN-2003; 2002US-0386447P.

XX 06-JUN-2003; 2002US-0386459P.

XX 06-JUN-2003; 2002US-0386465P.

XX 06-JUN-2003; 2002US-0386864P.

XX 07-JUN-2003; 2002US-0386701P.

XX 07-JUN-2003; 2002US-0386796P.

XX 07-JUN-2003; 2002US-0386931P.

XX 07-JUN-2003; 2002US-0387078P.

XX 07-JUN-2003; 2002US-0387081P.

XX 07-JUN-2003; 2002US-0387083P.

XX 10-JUN-2003; 2002US-0387429P.

XX 10-JUN-2003; 2002US-0387540P.

XX 10-JUN-2003; 2002US-0387866P.

XX 11-JUN-2003; 2002US-0387606P.

XX 11-JUN-2003; 2002US-0387610P.

XX 11-JUN-2003; 2002US-0387659P.

XX 11-JUN-2003; 2002US-0387668P.

XX 11-JUN-2003; 2002US-0387696P.

XX 11-JUN-2003; 2002US-0387859P.

PR 12-JUN-2002; 2002US-0387934P.  
PR 12-JUN-2002; 2002US-0387960P.  
PR 12-JUN-2002; 2002US-0388022P.  
PR 12-JUN-2002; 2002US-0388096P.  
PR 12-JUN-2002; 2002US-0388432P.  
PR 12-JUN-2002; 2002US-0388479P.  
PR 13-JUN-2002; 2002US-0389123P.  
PR 13-JUN-2002; 2002US-0389120P.  
PR 14-JUN-2002; 2002US-0389146P.  
PR 17-JUN-2002; 2002US-0389742P.  
PR 18-JUN-2002; 2002US-0389604P.  
PR 18-JUN-2002; 2002US-0389884P.  
PR 19-JUN-2002; 2002US-0390066P.  
PR 19-JUN-2002; 2002US-0390144P.  
PR 19-JUN-2002; 2002US-0390209P.  
PR 25-JUN-2002; 2002US-0391726P.  
PR 06-AUG-2002; 2002US-0401628P.  
PR 09-AUG-2002; 2002US-0402268P.  
PR 12-AUG-2002; 2002US-0402822P.  
PR 13-AUG-2002; 2002US-0403458P.  
PR 15-AUG-2002; 2002US-0403617P.  
PR 15-AUG-2002; 2002US-0403732P.  
PR 26-AUG-2002; 2002US-0406182P.  
PR 12-SEP-2002; 2002US-0410085P.  
PR 13-SEP-2002; 2002US-0410505P.  
PR 23-SEP-2002; 2002US-0412955P.  
PR 30-SEP-2002; 2002US-0415195P.  
PR 23-OCT-2002; 2002US-0420627P.  
PR 24-OCT-2002; 2002US-0420718P.  
PR 24-OCT-2002; 2002US-0420852P.  
PR 31-OCT-2002; 2002US-0422750P.  
PR 01-NOV-2002; 2002US-0423095P.  
PR 05-NOV-2002; 2002US-0423748P.  
XX (CURA-) CURAGEN CORP.  
XX  
XX Alsobrook JP, Anderson DM, Baumgartner JC, Berghs C, Boldog FL;  
PI Burgess CE, Casman SJ, Caterton E, Dhanabal M, Edinger SR;  
PI Ellerman K, Ettenberg S, Gangoli EA, Gerlach VL, Gorman L;  
PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;  
PI Khrantsov NV, Larocheville W, Li L, Liang H, Low K, Macdougall JR;  
PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CB;  
PI Millet I, Padigar M, Patturajan M, Peyman JA, Qian X, Rastelli L;  
PI Rieger DK, Rothenberg ME, Shenoy SG, Shinkets RA, Smithson G;  
PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CM, Voss EZ;  
PI Wolenc AR, Zhong M, Zhong H;  
XX WPI; 2004-053467/05.  
DR P-PSDB; ADH41621.  
XX  
XX New NOVX polypeptides and nucleic acid molecules useful for preventing or  
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in  
PT pharmacogenomics.  
XX  
PS Claim 20; SEQ ID NO 173; 1503pp; English.  
XX  
CC The invention relates to 566 new isolated human polypeptides and their  
CC encoding genes, sequences that are at least 95% identical to these or  
CC sequences comprising one or more conservative substitutions in these. The  
CC polypeptide, polynucleotide and antibodies against the polypeptides are  
CC useful in diagnosing, treating or preventing NOVX-associated disorders,  
CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,  
CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,  
CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.  
CC The nucleic acids are further used as hybridization probes, in chromosome  
CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The  
CC polypeptides are also useful as vaccines. This sequence represents an  
CC example of the nucleic acid sequence of the invention.  
XX  
SQ Sequence 2739 BP; 918 A; 550 C; 529 G; 742 T; 0 U; 0 Other;  
Query Match 38.2%; Score 2716.2; DB 12; Length 2739;

Best Local Similarity 99.7%; Pred. No. 0;  
Matches 2721; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 155 CACAGGTGATGTTTCCAAATGTCGTTCTCTGTTAGTACGATATACCATCTCTTCAA 214  
DB |||||  
5 CTCAGGTAGATGTTTCCAAATGTCGTTCTCTGTTAGTACGATATACCATCTCTTCAA 64  
QY 215 TTTCTACAAACATACACCTCACCTGTTACTAGAAATAGTGACACCAATGTAAACAAACAG 274  
DB |||||  
65 TTTCTACAAACATACACCTCACCTGTTACTAGAAATAGTGACACCAATGTAAACAAACAG 124  
QY 275 GGCCTCCAGTCTTCTCTAGCCGGGAAAGATCGGATCTCTGTTGGGATTTCTTGTCTTGA 334  
DB |||||  
125 GGCCTCCAGTCTTCTCTAGCCGGGAAAGATCGGATCTCTGTTGGGATTTCTTGTCTTGA 184  
QY 335 ATACACACCTTAATCCAAATGGAAGATATATCTTACATTTGTCAAATATAGGAAGTTT 394  
DB |||||  
185 ATACACACCTTAATCCAAATGGAAGATATATCTTACATTTGTCAAATATAGGAAGTTT 244  
QY 395 GTCCGTGGATGCAAAACAGTATATACAAAGTCAGATCAAAAGCCAGACAGTCTGGAAGTTC 454  
DB |||||  
245 GTCCGTGGATGCAAAACAGTATATACAAAGTCAGATCAAAAGCCAGACAGTCTGGAAGTTC 304  
QY 455 TTCTTACTAATCTTAATCTCTGGAACAACATATGAAATTAAGGTTGCTGCTGAAACAGTG 514  
DB |||||  
305 TTCTTACTAATCTTAATCTCTGGAACAACATATGAAATTAAGGTTGCTGCTGAAACAGTG 364  
QY 515 CTGGCATTTGGAGTGTGTTAGTGATCAATTTCTTCCAACTGCAGAAAGTGTCTCAGGAA 574  
DB |||||  
365 CTGGCATTTGGAGTGTGTTAGTGATCAATTTCTTCCAACTGCAGAAAGTGTCTCAGGAA 424  
QY 575 AAGTGGTGAATCTCAGTGTGAGGCTTACAGCTTCAACGCTTACAGCTTAAGCTGATTTGCTATT 634  
DB |||||  
425 AAGTGGTGAATCTCAGTGTGAGGCTTACAGCTTCAACGCTTACAGCTTAAGCTGATTTGCTATT 484  
QY 635 TACCTCGGCAACCAAAATGGCAAAATTTACCAGCTTCAAGATTTAGTGTCAAACTATGCCAGAA 694  
DB |||||  
485 TACCTCGGCAACCAAAATGGCAAAATTTACCAGCTTCAAGATTTAGTGTCAAGCATATGCCAGAA 544  
QY 695 GTGGATAGTAGTGAAAGATGTCATCAATCAGAGTAGAGACACATTTTGTAGTGGGAAATGTC 754  
DB |||||  
545 GTGGATAGTAGTGAAAGATGTCATCAATCAGAGTAGAGACACATTTTGTAGTGGGAAATGTC 604  
QY 755 CAGATGCAATGAGAAATAGTGAATCTTTTATGAGTAGACAGCCAGCCCTTCTCCAAACC 814  
DB |||||  
605 CAGATGCAATGAGAAATAGTGAATCTTTTATGAGTAGACAGCCAGCCCTTCTCCAAACC 664  
QY 815 TTGGTAGAGTTTACACCTCCATCGCGTACACACATTTCAATCAGCAGCTGTGACACAGAAATG 874  
DB |||||  
665 TTGGTAGAGTTTACACCTCCATCGCGTACACACATTTCAATCAGCAGCTGTGACACAGAAATG 724  
QY 875 AGATCAGCTCTGTGTGAAAGAGCTATCAGTTTGTAGTAGACACATTTGAGACCTTATA 934  
DB |||||  
725 AGATCAGCTCTGTGTGAAAGAGCTATCAGTTTGTAGTAGACACATTTGAGACCTTATA 784  
QY 935 CAAACATATCTTTTGAAGTTTTCAGCTGCTTACAACTGGAAGCAGGTTATATTAGTAGTACGA 994  
DB |||||  
785 CAAACATATCTTTTGAAGTTTTCAGCTGCTTACAACTGGAAGCAGGTTATATTAGTAGTACGA 844  
QY 995 TTGTGAGAAACACAGAAATCAGTGTCTGGAAGGACCAACCAAACTGCGTAAACAGGCAACA 1054  
DB |||||  
845 TTGTGAGAAACACAGAAATCAGTGTCTGGAAGGACCAACCAAACTGCGTAAACAGGCAACA 904  
QY 1055 TCACAGGAAAGTCTTTTCAATTTTATGGGACCCCAACACTATAGTAACAGGGAATTTA 1114  
DB |||||  
905 TCACAGGAAAGTCTTTTCAATTTTATGGGACCCCAACACTATAGTAACAGGGAATTTA 964  
QY 1115 GTTATAGAGTTGAAATATATATGCAACCATCAGGTGCGATTTTGGATTAACAGCAAAAGACC 1174  
DB |||||  
965 GTTATAGAGTTGAAATATATATGCAACCATCAGGTGCGATTTTGGATTAACAGCAAAAGACC 1024  
QY 1175 TCAAGTTTGCATTTCACTAACCTTAACACCATTTTACAAATGTATGATGTCTATATTGGCGTGT 1234  
DB |||||

Db 1025 TCAAGTTTGCATCTCACTAACCTAACCATTACCAATTTATGATGTCTATATATGCGGCTG 1084  
QY 1235 AAACCAAGTCAGGGAGCTGGGCCCAAGTCAAATATTTTCAGTATTCCTCCACCAAGATGTTTC 1294  
Db 1085 AAACCAAGTCAGGGAGCTGGGCCCAAGTCAAATATTTTCAGTATTCCTCCACCAAGATGTTTC 1144  
QY 1295 CAGGGCAGTGTGTTGATTTACAACTTGCAGAGTAGAATCCAGCAAGTAAGAATTTACTT 1354  
Db 1145 CAGGGCAGTGTGTTGATTTACAACTTGCAGAGTAGAATCCAGCAAGTAAGAATTTACTT 1204  
QY 1355 GGAAGAAACACCAACCAACAAATGGAATTTATTAACCAATACCGAGTGAAGTCTAGTTTC 1414  
Db 1205 GGAAGAAACACCAACCAACAAATGGAATTTATTAACCAATACCGAGTGAAGTCTAGTTTC 1264  
QY 1415 CAGAGACAGGAATTAATTTTGGAAATACTTTGCTCTACCTGGAATAATGAGTATATAAATG 1474  
Db 1265 CAGAGACAGGAATTAATTTTGGAAATACTTTGCTCTACCTGGAATAATGAGTATATAAATG 1324  
QY 1475 ACCCCATGGCTCAGAAATTTGTGAACATAGTAGAGCAATGGTAGGATTTATAGAGGTT 1534  
Db 1325 ACCCCATGGCTCAGAAATTTGTGAACATAGTAGAGCAATGGTAGGATTTATAGAGGTT 1384  
QY 1535 CAGCAGAGATGTCTGTGACCTTTCCTCTACCTCTACCTTATATATAACAGCATCCAG 1594  
Db 1385 CAGCAGAGATGTCTGTGACCTTTCCTCTACCTCTACCTTATATATAACAGCATCCAG 1444  
QY 1595 ATAAAACTTTCTGCAAGGAATAGAGCTGAAGACAGAGCTTCCACAGTTGTAACTACAA 1654  
Db 1445 ATAAAACTTTCTGCAAGGAATAGAGCTGAAGACAGAGCTTCCACAGTTGTAACTACAA 1504  
QY 1655 GGAATCAGTATATTAAGTATGAGTGAACAGAGCTGTCTTATGTTATCAGGAGACTTG 1714  
Db 1505 GGAATCAGTATATTAAGTATGAGTGAACAGAGCTGTCTTATGTTATCAGGAGACTTG 1564  
QY 1715 TACCTTTCTAGTACACATGATTTAGTGTCTCTCTTCCATCATGGGAGGAGGACAC 1774  
Db 1565 TACCTTTCTAGTACACATGATTTAGTGTCTCTCTTCCATCATGGGAGGAGGACAC 1624  
QY 1775 CAACAGTTCTCAGTGTAGGACACGTCAGCAAGTGCCAAAGCTCCATTAATAATTTAACT 1834  
Db 1625 CAACAGTTCTCAGTGTAGGACACGTCAGCAAGTGCCAAAGCTCCATTAATAATTTAACT 1684  
QY 1835 ATAAAAATATAGTCTTCACTATTTTGTATATTTGGGATCTCCAGAAATATCCCAATG 1894  
Db 1685 ATAAAAATATAGTCTTCACTATTTTGTATATTTGGGATCTCCAGAAATATCCCAATG 1744  
QY 1895 GAAAAATACCTACTATAGATTTATGCAATGGAATTTGGATACAAACAGAGCAATCCAGA 1954  
Db 1745 GAAAAATACCTACTATAGATTTATGCAATGGAATTTGGATACAAACAGAGCAATCCAGA 1804  
QY 1955 TAACTACCATAGATAACAGCTTTCTCATACAGGGTTTAAAGAAATACAAAAATACAAA 2014  
Db 1805 TAACTACCATAGATAACAGCTTTCTCATACAGGGTTTAAAGAAATACAAAAATACAAA 1864  
QY 2015 TGAGATGGCAGCTCAACCCAGATGGAGAAAGTCTTTGCTGTAAGAAATATGACATCT 2074  
Db 1865 TGAGATGGCAGCTCAACCCAGATGGAGAAAGTCTTTGCTGTAAGAAATATGACATCT 1924  
QY 2075 TTGTGAGAACTTCAGAGATGAACCGGAATCATCCTCAAGATGTGGAAGTAATTTGATG 2134  
Db 1925 TTGTGAGAACTTCAGAGATGAACCGGAATCATCCTCAAGATGTGGAAGTAATTTGATG 1984  
QY 2135 TTACCCGAGATGAATAAGTGTGAAGTGTGTCACACCGGAAAGCCCAATGGGATCATTA 2194  
Db 1985 TTACCCGAGATGAATAAGTGTGAAGTGTGTCACACCGGAAAGCCCAATGGGATCATTA 2044  
QY 2195 TTGCTTATGAAGTGTATATAAAAAATAGATATCTTTATATATGAAGAAACACATCAACAA 2254  
Db 2045 TTGCTTATGAAGTGTATATAAAAAATAGATATCTTTATATATGAAGAAACACATCAACAA 2104  
QY 2255 CAGACATATATTAAGAACTTAAGACCTTCAACCCCTCTATATACATTTCTGTAAGTCTT 2314  
Db 2105 CAGACATATATTAAGAACTTAAGACCTTCAACCCCTCTATATACATTTCTGTAAGTCTT 2164

QY 2315 ACAACAGATTTGGTCATGGCAATCAGGTATCTTTTACTCTCTGTGAAGACTTCGGAGA 2374  
Db 2165 ACAACAGATTTGGTCATGGCAATCAGGTATCTTTTACTCTCTGTGAAGACTTCGGAGA 2224  
QY 2375 CTGTGCTGATAGTCGACGAGAAATATACATCTACAAATATTTCTTCTGGAGAGATTG 2434  
Db 2225 CTGTGCTGATAGTCGACGAGAAATATACATCTACAAATATTTCTTCTGGAGAGATTG 2284  
QY 2435 AGCTATCATTTCTTCCCCCAAGTAGTCCCAATGGAATCATAAAAAATATACAAATTTATC 2494  
Db 2285 AGCTATCATTTCTTCCCCCAAGTAGTCCCAATGGAATCATAAAAAATATACAAATTTATC 2344  
QY 2495 TCAAGAGAAATGTAAGAAATGAGAAAGAACTATATAAATCAACCTCTTTAAACCCAAAAACA 2554  
Db 2345 TCAAGAGAAATGTAAGAAATGAGAAAGAACTATATAAATCAACCTCTTTAAACCCAAAAACA 2404  
QY 2555 TTAAAGTACTGAAGAAATATACCCCAATATATCATTTGAGGTGTCTGTAGTACACTGAAAG 2614  
Db 2405 TTAAAGTACTGAAGAAATATACCCCAATATATCATTTGAGGTGTCTGTAGTACACTGAAAG 2464  
QY 2615 GTGAAGGAGTTCCGAGTGTCTCCATAAAGTATATCTACGAGGAGGAAGATGCTCTGATTTCTC 2674  
Db 2465 GTGAAGGAGTTCCGAGTGTCTCCATAAAGTATATCTACGAGGAGGAAGATGCTCTGATTTCTC 2524  
QY 2675 CCCCTCAAGACTTCTGTGTAAACAGTTCTGTGTGTCTACGGTGAAGTTGTCTANGCAAC 2734  
Db 2525 CCCCTCAAGACTTCTGTGTAAACAGTTCTGTGTGTCTACGGTGAAGTTGTCTANGCAAC 2584  
QY 2735 CACCCCTGGAGCAAAATGGAATTTATTTATACACAGTTTATGTCTCGAATAGATCAT 2794  
Db 2585 CACCCCTGGAGCAAAATGGAATTTATTTATTTATACACAGTTTATGTCTCGAATAGATCAT 2644  
QY 2795 CATTAAAACTATTAATGTCTACGAAACATCATTTGGAGTTATCAGATTTGGATTATAATG 2854  
Db 2645 CATTAAAACTATTAATGTCTACGAAACATCATTTGGAGTTATCAGATTTGGATTATAATG 2704  
QY 2855 TTGAATACAGTCTTATGTAAACAGCTAGC 2883  
Db 2705 TTGAATACAGTCTTATGTAAACAGCTAGC 2733

## RESULT 12

ABT06282

ID ABT06282 standard; cDNA; 2565 BP.

XX AC ABT06282;

XX DT 24-OCT-2002 (first entry)

XX DE Human NOV2b coding sequence.

Human; NOVX; autoimmune disease; cancer; infection; inflammatory disease;  
storage disorder; muscle disorder; neurodegenerative disorder; hypotensive;  
developmental defect; neuroprotective; antiparkinsonian; hypotensive;  
hypertensive; haemostatic; cardiant; antilanginal; dermatological;  
immunosuppressive; antiinflammatory; virucide; antibacterial; anti-HIV;  
antiparasitic; antiallergic; antiasthmatic; antirheumatic; antiarthritic;  
vulnerary; anorectic; antidiabetic; immunomodulator; antipsoriatic;  
nephrotropic; kerolytic; antilucer; cerebroprotective; anticonvulsant;  
antinfertility; antimanic; antidepressant; metabolic; cycostatic;  
tranquillizer; analgesic; gene; ss..

OS Homo sapiens.

XX WO200257450-A2.

XX PD 25-JUL-2002.

XX PF 29-NOV-2001; 2001WO-US048922.

XX XX 29-NOV-2000; 2000US-0253834P.

PR 30-NOV-2000; 2000US-0250926P.



PR 25-JAN-2001; 2001US-0264180P.  
PR 20-AUG-2001; 2001US-0313656P.  
PR 05-OCT-2001; 2001US-0327456P.  
PR 28-NOV-2001; 2001US-00327456.  
XX (CURA-) CURAGEN CORP.  
XX  
XX Edinger S, Macdougall JR, Millet J, Ellerman K, Stone DJ;  
PI Gerlach V, Grosse WM, Alsbrook JP, Lepley DM, Rieger D, Burgess CE;  
PI Casman SJ, Spytek KA, Boldog FL, Li L, Padigar M, Mishra V;  
PI Patturajan M, Shenoy S, Rastelli L, Tchernev VT, Vernet CAM;  
PI Zerhusen BD, Malyankar UM, Guo X, Miller CE, Gangolli EA;  
XX  
XX WPI; 2002-590741/63.  
DR P-PSDB; AAO18737.  
XX  
XX Novel isolated polypeptide, designated NOVX, useful for treating or  
PT preventing in NOVX-associated disorders e.g. cardiomyopathy,  
PT atherosclerosis, diabetes, cancer, allergy, asthma, Crohn's disease.  
XX  
XX Claim 9; Page 26-27; 353pp; English.  
XX  
XX The present invention provides the protein and coding sequences of  
CC several novel human proteins, designated NOVX. These can be used in the  
CC treatment of, amongst others, cancers, autoimmune diseases, infections,  
CC inflammatory diseases, storage disorders, muscle disorders,  
CC neurodegenerative diseases and developmental defects. The present  
CC sequence is a coding sequence of the invention  
XX  
SQ Sequence 2565 BP; 885 A; 523 C; 447 G; 710 T; 0 U; 0 Other;  
  
Query Match 35.9%; Score 2551; DB 6; Length 2565;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2554; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
  
QY 3262 CCTGAAGGGTTGTTGGAAACCTGACTTACGAATCCATTTGGTCAACTGCAATTAATGTA 3321  
DB 7 CCTGAAGGGTTGTTGGAAACCTGACTTACGAATCCATTTGGTCAACTGCAATTAATGTA 66  
  
QY 3322 AGCTGGGTCCACCGGCTCAACAAACGGTCTAGTCTTCTATATGATGTTTCACTGATCTTA 3381  
DB 67 AGCTGGGTCCACCGGCTCAACAAACGGTCTAGTCTTCTATATGATGTTTCACTGATCTTA 126  
  
QY 3382 CAGCAGACTCCTCGCCATGTGAGACCCTCTGTTTACATATGAGAGAGCATATATTT 3441  
DB 127 CAGCAGACTCCTCGCCATGTGAGACCCTCTGTTTACATATGAGAGAGCATATATTT 186  
  
QY 3442 GATAATCTGGAAAAATACACTGATTATATATTAATAAATTAATCTCCATCAACAGAAAGGGA 3501  
DB 187 GATAATCTGGAAAAATACACTGATTATATATTAATAAATTAATCTCCATCAACAGAAAGGGA 246  
  
QY 3502 TTCTCTGATACCTTATCTGCCAGCTTATACATCAAGACTGAAGAAGATGCCAGAAACT 3561  
DB 247 TTCTCTGATACCTTATCTGCCAGCTTATACATCAAGACTGAAGAAGATGCCAGAAACT 306  
  
QY 3562 TCACCAATATCAACTTTTAAAAACCTTCTCTACCTCAGTCTTCTTATCATGGAT 3621  
DB 307 TCACCAATATCAACTTTTAAAAACCTTCTCTACCTCAGTCTTCTTATCATGGAT 366  
  
QY 3622 CCCCCAGTAAGCCAAATGGTGAATAATAAGTTATGATTTAACTTTTACAAGGACCAAAAT 3681  
DB 367 CCCCCAGTAAGCCAAATGGTGAATAATAAGTTATGATTTAACTTTTACAAGGACCAAAAT 426  
  
QY 3682 GAAAAATATCTTTTCAATTAATCTGATTAATAATACATAATATGGAAGAGCTTTTCAACATTT 3741  
DB 427 GAAAAATATCTTTTCAATTAATCTGATTAATAATATGGAAGAGCTTTTCAACATTT 486  
  
QY 3742 ACATTATATAGCTTTTTCGTCGCCGAAGCACTAGAAAGGACTTGGTCTTCCAGTATT 3801  
DB 487 ACATTATATAGCTTTTTCGTCGCCGAAGCACTAGAAAGGACTTGGTCTTCCAGTATT 546  
  
QY 3802 CTTTCTTTTACACAGATGAGTCCGCTTACACCTCCACAAATTTTGAATTTTAAATC 3861  
DB CTTTCTTTTACACAGATGAGTCCGCTTACACCTCCACAAATTTTGAATTTTAAATC 1686

DB 547 CTTTCTTTTACACAGATGAGTCCGCTTACACCTCCACAAATTTTGAATTTTAAATC 606  
QY 3862 AACTGTACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGATTT 3921  
DB 607 AACTGTACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGATTT 666  
QY 3922 GTTAAAGTATATAGTTTAAAAATTCATGAACATGAATGAACATGACATATATATTAAGAAT 3981  
DB 667 GTTAAAGTATATAGTTTAAAAATTCATGAACATGAATGAACATGACATATATATTAAGAAT 726  
QY 3982 ATATCAGGATTTAAACTGAAGCCAAACTTTGTTGGACTGGAAACCAAGTCAAGCTACTCT 4041  
DB 727 ATATCAGGATTTAAACTGAAGCCAAACTTTGTTGGACTGGAAACCAAGTCAAGCTACTCT 786  
QY 4042 ATCCGTGTATCTCGGTTCCAAAGTTGGAATGGAATCAATTTAGTAAATAGTAAATA 4101  
DB 787 ATCCGTGTATCTCGGTTCCAAAGTTGGAATGGAATCAATTTAGTAAATAGTAAATA 846  
QY 4102 TTCACAACCCCAAGAAATCAGTTCAGATGTCGTGCAAGATATATGAGTGCATGCAACTAGC 4161  
DB 847 TTCACAACCCCAAGAAATCAGTTCAGATGTCGTGCAAGATATATGAGTGCATGCAACTAGC 906  
QY 4162 TGGCAGTCAAGTTTATGTAATGGATCCACCCAAAGGCAAAATGGAATATAACGCAG 4221  
DB 907 TGGCAGTCAAGTTTATGTAATGGATCCACCCAAAGGCAAAATGGAATATAACGCAG 966  
QY 4222 TATATGTTAAAGTTGGAAGGAAATTTCTACAAAGTTTCTCCCAAGATCAGATGATACACT 4281  
DB 967 TATATGTTAAAGTTGGAAGGAAATTTCTACAAAGTTTCTCCCAAGATCAGATGATACACT 1026  
QY 4282 TTCTATAAGCTTTCTTGGCAATACCTCATATGTCCTTTAAAGTAAAGCTTTCAACCTCAGCT 4341  
DB 1027 TTCTATAAGCTTTCTTGGCAATACCTCATATGTCCTTTAAAGTAAAGCTTTCAACCTCAGCT 1086  
QY 4342 GGTGAAGGTGATGAAAGCACTGCCATGTGAGCACTACTCTGAAACAGTTCCAGTGT 4401  
DB 1087 GGTGAAGGTGATGAAAGCACTGCCATGTGAGCACTACTCTGAAACAGTTCCAGTGT 1146  
QY 4402 CCCACAAATATGCTTTTCTGATGTTTCTGATGTTTCTGATGTTTCTGATGTTTCTGATGTTT 4461  
DB 1147 CCCACAAATATGCTTTTCTGATGTTTCTGATGTTTCTGATGTTTCTGATGTTTCTGATGTTT 1206  
QY 4462 CCTGACACTATCTTGGCTTACTTTTCAAAATTAACAAATTAACCACTCAACTCTGCTCAA 4521  
DB 1207 CCTGACACTATCTTGGCTTACTTTTCAAAATTAACAAATTAACCACTCAACTCTGCTCAA 1266  
QY 4522 AAATGCAAGAAATGGGAATCCGAAGATGTGTTGAATATCAAAATTAACCACTCAACTCTGCTCAA 4581  
DB 1267 AAATGCAAGAAATGGGAATCCGAAGATGTGTTGAATATCAAAATTAACCACTCAACTCTGCTCAA 1326  
QY 4582 GAAGCTCACTTAACCTGAGAGAGAGATATATGGAATTAAGAAATTTAGATGTTATAGATTC 4641  
DB 1327 GAAGCTCACTTAACCTGAGAGAGAGATATATGGAATTAAGAAATTTAGATGTTATAGATTC 1386  
QY 4642 CAAAGTGGCTCCAGCAGCACTGCTGCTATGCAATGCTTCAAACTGGAATTTCTACAAA 4701  
DB 1387 CAAAGTGGCTCCAGCAGCACTGCTGCTATGCAATGCTTCAAACTGGAATTTCTACAAA 1446  
QY 4702 ACTCTGCTGGCCCTCCAGATGGTCTCTGAAAAATGTTTCATGTAGTAGCAACATCACCT 4761  
DB 1447 ACTCTGCTGGCCCTCCAGATGGTCTCTGAAAAATGTTTCATGTAGTAGCAACATCACCT 1506  
QY 4762 TTTAGCATCAGATTAAGCTGGAGTGAACCTGCTGTCATTTACCTGGACCAACATGTTATCTG 4821  
DB 1507 TTTAGCATCAGATTAAGCTGGAGTGAACCTGCTGTCATTTACCTGGACCAACATGTTATCTG 1566  
QY 4822 ATTGATGTCAAAATCGGTAGATTAATGATGAATTTTAAATATATATCTCATCAAGTCAATGAA 4881  
DB 1567 ATTGATGTCAAAATCGGTAGATTAATGATGAATTTTAAATATATCTCATCAAGTCAATGAA 1626  
QY 4882 GAAAAATAAACCAATGAAGAAATTAAGAAATTTAGAAATTTTCAAGGTTATTTCTGTAGTGATC 4941  
DB 1627 GAAAAATAAACCAATGAAGAAATTTAGAAATTTTCAAGGTTATTTCTGTAGTGATC 1686









```
QY 2458 AGTCCCAATGGAATCATATAAAATATATACAAATTTATCTCAAGAGAAATGTAATGGAATGAG 2517
Db 1447 AGTCCCAATGGAATCATATAAAATATATACAAATTTATCTCAAGAGAAATGTAATGGAATGAG 1506
QY 2518 GAAAGAACTATAATATACAACTCTTTAAACCCAAACATTAAGTACTTGAAGAATATATACC 2577
Db 1507 GAAAGAACTATAATATACAACTCTTTAAACCCAAACATTAAGTACTTGAAGAATATATACC 1566
QY 2578 CAATATATCATTCAGGTGTCGTCTAGTACACTGAAAGGTGAAGGAGTTCGGAGTCTCC 2637
Db 1567 CAATATATCATTCAGGTGTCGTCTAGTACACTGAAAGGTGAAGGAGTTCGGAGTCTCC 1626
QY 2638 ATAAGTATATCTAGCGAGGAGATGCTCTCTGATTTCTCCCTCAAGACTTCTCTGTAATA 2697
Db 1627 ATAAGTATATCTAGCGAGGAGATGCTCTCTGATTTCTCCCTCAAGACTTCTCTGTAATA 1686
QY 2698 CAGTGTGCTGCTGACGGTGAAGTGTCTATGCAACCAACCCCTGAGGCAAAATGGAAT 2757
Db 1687 CAGTGTGCTGCTGACGGTGAAGTGTCTATGCAACCAACCCCTGAGGCAAAATGGAAT 1746
QY 2758 ATCTTTTATPACACAGTTTATGTCGGAATAGATCATATTAATAAACTATTAATGTCAC 2817
Db 1747 ATCTTTTATPACACAGTTTATGTCGGAATAGATCATATTAATAAACTATTAATGTCAC 1806
QY 2818 GAAACATCATTTGAGTTATCAGATTTGGATTAATGTTGAATACAGTCTTATGTAACA 2877
Db 1807 GAAACATCATTTGAGTTATCAGATTTGGATTAATGTTGAATACAGTCTTATGTAACA 1866
QY 2878 GCTAGCACCAAGATTTGCTGATGCGGAACAGGAGCAATATCATTAAGTCTTCAACACCA 2937
Db 1867 GCTAGCACCAAGATTTGCTGATGCGGAACAGGAGCAATATCATTAAGTCTTCAACACCA 1926
QY 2938 GAGGAGCACCAGCAGTCTCCCAAGATGTTTATATGCAAACTCTGATCTTCTATCA 2997
Db 1927 GAGGAGCACCAGCAGTCTCCCAAGATGTTTATATGCAAACTCTGATCTTCTATCA 1986
QY 2998 ATAAATCTTTCTGGAACCTCTCTCAAACTTAATGGAATATACAAATTAATCTCTGTT 3057
Db 1987 ATAAATCTTTCTGGAACCTCTCTCAAACTTAATGGAATATACAAATTAATCTCTGTT 2046
QY 3058 TATTACAGAAATCTTCAGTACTTTTATGCAAACTTATGCAAACTTATGCAAACTTATGCA 3117
Db 2047 TATTACAGAAATCTTCAGTACTTTTATGCAAACTTATGCAAACTTATGCAAACTTATGCA 2106
QY 3118 GACTTTGACAATATGACTGTATCCAAATTTATAGATAAACTGCAATATTTCACTACTAT 3177
Db 2107 GACTTTGACAATATGACTGTATCCCAATTTATAGATAAACTGCAATATTTCACTACTAT 2166
QY 3178 ACATTTGGTTAAACAGCAAGTACTTCAGTTGGAAATGGGAATAAAGCAGTGA 3230
Db 2167 ACATTTGGTTAAACAGCAAGTACTTCAGTTGGAAATGGGAATAAAGCCTCGA 2219
```

## RESULT 15

AAQ73786

ID AAQ73786 standard; cDNA; 2309 BP.

XX AC AAQ73786;

XX 25-MAR-2003 (revised)

DT 26-JUN-1995 (first entry)

XX Partial PCR fragment of PTP-S31D.

DE Protein tyrosine phosphatase; cancer; diabetes; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 65..1035

FT misc\_feature 1766

/\*tag= b  
/note= "unknown base"

W09421800-A2.

29-SEP-1994.

23-MAR-1994; 94WO-BP000909.

23-MAR-1993; \*93US-00036210.

(PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

Moller NPH, Moller KB, Ullrich A;

WPI; 1994-317020/39.

P-PSDB; AAR60877.

New protein tyrosine phosphatase PTP-S31 - is used to develop prods. for

treating or preventing disease associated with abnormal PTP-S31, e.g.

cancer or diabetes.

Claim 6; Fig 6; 116pp; English.

The sequence is that of a partial fragment of a new protein tyrosine

phosphatase PTP-S31D that was isolated from RNA from human skeletal

muscle. Such DNA can be used to treat cancer and diabetes. See also

AAQ73782-8. (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 2309 BP; 770 A; 384 C; 435 G; 719 T; 0 U; 1 Other;

Query Match 15.1%; Score 1071.6; DB 2; Length 2309;

Best Local Similarity 99.4%; Pred. No. 1.4e-245;

Matches 1087; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 6015 AACGGGCAATTAAGCAAGAAATCTCTGCAATGTTGAGAGCTTTGCACAAACAA 6074

Db 70 AATGAGGCAATTAAGCAAGAAATCTCTGCAATGTTGAGAGCTTTGCACAAACAA 129

QY 6075 CAACCTAAAGTTTCAAGAAAGAAATTTTCCGAATTTACCAAAATTTCTTCAGGATCTTTCTTC 6134

Db 130 CAACCTAAAGTTTCAAGAAAGAAATTTTCCGAATTTACCAAAATTTCTTCAGGATCTTTCTTC 189

QY 6135 AACTGATGCTGATCTGCTTGAATAGAGCAAAACCGTTTCCCAAAACATAAAACATA 6194

Db 190 AACTGATGCTGATCTGCTTGAATAGAGCAAAACCGTTTCCCAAAACATAAAACATA 248

QY 6195 TAATTAATATACAGAGTAAAGCTGATAGCTGACGCTAGTGTTCAGGTTCCGATTTATAT 6254

Db 249 --ATAATAATAACAGAGTAAAGCTGATAGCTGACGCTAGTGTTCAGGTTCCGATTTATAT 306

QY 6255 TAATGCCAGCTATATTTCTGGTTATTTATGTCAAATGAATTTATTTGCTACTCAAGGTC 6314

Db 307 TAATGCCAGCTATATTTCTGGTTATTTATGTCAAATGAATTTATTTGCTACTCAAGGTC 366

QY 6315 ACTACCAAGCAAGTGGAGATTTTGGAGAAATGTTGGGAAACCCAGGCAAAACATTT 6374

Db 367 ACTACCAAGCAAGTGGAGATTTTGGAGAAATGTTGGGAAACCCAGGCAAAACATTT 426

QY 6375 AGTAATGCTAACACAGTGTTTTGAAGAGGAGGATCAGATGCCATCAGTATTTGGCCAGA 6434

Db 427 AGTAATGCTAACACAGTGTTTTGAAGAGGAGGATCAGATGCCATCAGTATTTGGCCAGA 486

QY 6435 GGACCAACCAAGCCAGTTACTGCTTTGGAGATATAGTATTACAAAGCTAATGAGGATGT 6494

Db 487 GGACCAACCAAGCCAGTTACTGCTTTGGAGATATAGTATTACAAAGCTAATGAGGATGT 546

QY 6495 TCAATAGATTGGACTATCAGGGATCTGAAAATGGAAGGCAATGGGATTTGCATGCTGT 6554

Db 547 TCAATAGATTGGACTATCAGGGATCTGAAAATGGAAGGCAATGGGATTTGCATGCTGT 606

QY 6555 TCGACAGTGTAACTTTACTGCTGCTGCCAGAGATGGGTTTCTGAGAACAGGCCCTCT 6614

Db 607 TCGACAGTGTAACTTTACTGCTGGCCAGAGCATGGGTTCTGTGAGAACAGCGCCCTCT 666  
Qy 6615 AATTCACTTTGGAAGTTGGTTGCGAGCAAGCAGGGCACAATGACACACCACTATGATTGT 6674  
Db 667 AATTCACTTTGGAAGTTGGTTGCGAGCAAGCAGGGCACAATGACACCACTATGATTGT 726  
Qy 6675 TCACTGCAGTGTGGAAGTTGGAAGAACTGAGTGTATTTATTTGCTCTGGACCAATTTAACACA 6734  
Db 727 TCACTGCAGTGTGGAAGTTGGAAGAACTGAGTGTATTTATTTGCTCTGGACCAATTTAACACA 786  
Qy 6735 ACATATAAATGACCAATGATTTTGTGGATATATATGACTAGTAGCTGAACTGAGAAGTGA 6794  
Db 787 ACATATAAATGACCAATGATTTTGTGGATATATATGACTAGTAGCTGAACTGAGAAGTGA 846  
Qy 6795 AGAATGTGATGGTGCAGAACTGCGCAGATATATCTTTTACACCAGTGCATTCTGGA 6854  
Db 847 AAGAATGTGATGGTGCAGAACTGCGCAGATATATCTTTTACACCAGTGCATTCTGGA 906  
Qy 6855 TCTCTTATCAAAATAAGGGAAGTAATCAGCCCATCTGTTTGTAACTATTACAGCACTTCA 6914  
Db 907 TCTCTTATCAAAATAAGGGAAGTAATCAGCCCATCTGTTTGTAACTATTACAGCACTTCA 966  
Qy 6915 GAAGATGGACTCTTTGGACGCCATGGAAGGTGATTTGAGCTTGAATGGGAAGAAACCAC 6974  
Db 967 GAAGATGGACTCTTTGGACGCCATGGAAGGTGATTTGAGCTTGAATGGGAAGAAACCAC 1026  
Qy 6975 TATGTAATAATTCAGACCAAGGATACAAATGGAAGAGATTTTAAATCCAGGGGCCAA 7034  
Db 1027 TATGTAATAATTCAGACCAAGGATACAAATGGAAGAGATTTTAAATCCAGGGGCCAA 1086  
Qy 7035 AGTTACCCCTCATCTTCCGAATTCGAATGCAACCTTAAAGAAATATCTATGCTTCT 7094  
Db 1087 AGTTACCCCTCATCTTCCGAATTCGAATGCAACCTTAAAGAAATATCTATGCTTCT 1146  
Qy 7095 CTCACCTGTGCTTT 7108  
Db 1147 CTCACCTGTGCTTT 1160

Search completed: July 1, 2005, 22:46:26  
Job time : 5236 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2005, 21:03:34 ; Search time 1608 Seconds  
(without alignments)  
7232.994 Million cell updates/sec

Title: US-10-673-885-1  
Perfect score: 7108  
Sequence: 1 taattgtactgtccagaa.....gcttctctactgtgccttt 7108

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description        |
|------------|--------|-------------|--------|-------|--------------------|
| 1          | 7108   | 100.0       | 7108   | 4     | US-09-822-871-1    |
| 2          | 2618.6 | 36.8        | 3973   | 1     | US-08-036-210-21   |
| 3          | 2618.6 | 36.8        | 3973   | 2     | US-08-449-609-21   |
| 4          | 2618.6 | 36.8        | 3973   | 4     | US-09-361-096A-21  |
| 5          | 1118.8 | 15.7        | 2692   | 1     | US-08-036-210-14   |
| 6          | 1118.8 | 15.7        | 2692   | 2     | US-08-449-609-14   |
| 7          | 1118.8 | 15.7        | 2692   | 4     | US-09-361-096A-14  |
| 8          | 1071.6 | 15.1        | 2309   | 1     | US-08-036-210-10   |
| 9          | 1071.6 | 15.1        | 2309   | 2     | US-08-449-609-10   |
| 10         | 1071.6 | 15.1        | 2309   | 4     | US-09-361-096A-10  |
| 11         | 789.6  | 11.1        | 2173   | 1     | US-08-036-210-6    |
| 12         | 789.6  | 11.1        | 2173   | 2     | US-08-449-609-6    |
| 13         | 789.6  | 11.1        | 2173   | 4     | US-09-361-096A-6   |
| 14         | 532    | 7.5         | 532    | 1     | US-08-036-210-8    |
| 15         | 532    | 7.5         | 532    | 2     | US-08-449-609-8    |
| 16         | 532    | 7.5         | 532    | 4     | US-09-361-096A-8   |
| 17         | 319.4  | 4.5         | 321    | 1     | US-08-036-210-3    |
| 18         | 319.4  | 4.5         | 321    | 2     | US-08-449-609-3    |
| 19         | 319.4  | 4.5         | 321    | 4     | US-09-361-096A-3   |
| 20         | 315    | 4.4         | 254366 | 4     | US-09-822-871-3    |
| 21         | 195.2  | 2.7         | 5679   | 1     | US-08-201-697-1    |
| 22         | 192.2  | 2.7         | 4064   | 4     | US-09-949-016-2962 |
| 23         | 192.2  | 2.7         | 4064   | 4     | US-09-949-016-2963 |
| 24         | 192.2  | 2.7         | 4064   | 4     | US-09-949-016-2964 |
| 25         | 192.2  | 2.7         | 4064   | 4     | US-09-949-016-2965 |
| 26         | 192.2  | 2.7         | 4064   | 4     | US-09-949-016-2966 |
| 27         | 192.2  | 2.7         | 4064   | 4     | US-09-949-016-2967 |

|    |       |     |      |   |                    |                   |
|----|-------|-----|------|---|--------------------|-------------------|
| 28 | 192.2 | 2.7 | 4752 | 1 | US-08-201-697-3    | Sequence 3, Appli |
| 29 | 192.2 | 2.7 | 4815 | 1 | US-08-201-697-5    | Sequence 5, Appli |
| 30 | 192.2 | 2.7 | 5069 | 4 | US-09-949-016-2180 | Sequence 2180, Ap |
| 31 | 192.2 | 2.7 | 5069 | 4 | US-09-949-016-2181 | Sequence 2181, Ap |
| 32 | 192.2 | 2.7 | 5069 | 4 | US-09-949-016-2182 | Sequence 2182, Ap |
| 33 | 192.2 | 2.7 | 5069 | 4 | US-09-949-016-2183 | Sequence 2183, Ap |
| 34 | 192.2 | 2.7 | 5069 | 4 | US-09-949-016-2184 | Sequence 2184, Ap |
| 35 | 192.2 | 2.7 | 5153 | 4 | US-09-949-016-2957 | Sequence 2957, Ap |
| 36 | 192.2 | 2.7 | 5153 | 4 | US-09-949-016-2958 | Sequence 2958, Ap |
| 37 | 192.2 | 2.7 | 5153 | 4 | US-09-949-016-2959 | Sequence 2959, Ap |
| 38 | 192.2 | 2.7 | 5153 | 4 | US-09-949-016-2960 | Sequence 2960, Ap |
| 39 | 192.2 | 2.7 | 5153 | 4 | US-09-949-016-2961 | Sequence 2961, Ap |
| 40 | 192.2 | 2.7 | 6075 | 4 | US-09-023-655-1366 | Sequence 1366, Ap |
| 41 | 192.2 | 2.7 | 6075 | 4 | US-09-949-016-404  | Sequence 404, App |
| 42 | 191.6 | 2.7 | 5763 | 4 | US-09-949-016-2178 | Sequence 2178, Ap |
| 43 | 187.4 | 2.6 | 4799 | 1 | US-08-201-697-6    | Sequence 6, Appli |
| 44 | 171.8 | 2.4 | 4756 | 4 | US-09-023-655-1461 | Sequence 1461, Ap |
| 45 | 171.8 | 2.4 | 5117 | 3 | US-08-854-585-1    | Sequence 1, Appli |

ALIGNMENTS

RESULT 1  
US-09-822-871-1  
; Sequence 1, Application US/09822871  
; Patent No. 6723547  
; GENERAL INFORMATION:  
; APPLICANT: WEBSTER, Marion et al  
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,  
; FILE REFERENCE: CL001219  
; CURRENT APPLICATION NUMBER: US/09/822,871  
; CURRENT FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 7108  
; TYPE: DNA  
; ORGANISM: Human  
US-09-822-871-1

|                       |                 |  |           |              |
|-----------------------|-----------------|--|-----------|--------------|
| Query Match           | 100.0%          | Score 7108;  | DB 4;     | Length 7108; |
| Best Local Similarity | 100.0%;         | Pred. No. 0;   |           |              |
| Matches 7108;         | Conservative 0; | Mismatches 0;  | Indels 0; | Gaps 0;      |
| QY                    | 1               | TAATTGTGTACTTGCAGAGGATCTGCTTTAAATCATTAATGCAGGCAACATTTCTCT    | 60        |              |
| Db                    | 1               | TAATTGTGTACTTGCAGAGGATCTGCTTTAAATCATTAATGCAGGCAACATTTCTCT    | 60        |              |
| QY                    | 61              | CTAGAGCCATCAATGTGATCTTCTGGCTGAAATAATGTAATAAGATGATTTCTTATC    | 120       |              |
| Db                    | 61              | CTAGAGCCATCAATGTGATCTTCTGGCTGAAATAATGTAATAAGATGATTTCTTATC    | 120       |              |
| QY                    | 121             | ATTTTCTTTTACTTTTATTTGGGACTTCAGAGACACAGGTTGATGTTTCCAATGTCGTT  | 180       |              |
| Db                    | 121             | ATTTTCTTTTACTTTTATTTGGGACTTCAGAGACACAGGTTGATGTTTCCAATGTCGTT  | 180       |              |
| QY                    | 181             | CTGTGTACTAGTAGATATAACCATCTTCAATTTCTAACACATACACCTCCTGTT       | 240       |              |
| Db                    | 181             | CTGTGTACTAGTAGATATAACCATCTTCAATTTCTAACACATACACCTCCTGTT       | 240       |              |
| QY                    | 241             | ACTAGATAGTACACCAATGTAACAAACAGGCTCCAGTCTTCTTAGCGGGGAA         | 300       |              |
| Db                    | 241             | ACTAGATAGTACACCAATGTAACAAACAGGCTCCAGTCTTCTTAGCGGGGAA         | 300       |              |
| QY                    | 301             | ACAGTCCGATCTGCTGGGATTTCTTGTCTTGGATATACACCACTTAATCAATGGAAG    | 360       |              |
| Db                    | 301             | ACAGTCCGATCTGCTGGGATTTCTTGTCTTGGATATACACCACTTAATCAATGGAAG    | 360       |              |
| QY                    | 361             | ATTATATCTTACATTTGTCAATATATAGGAAGTTTGTCCGTGGATGCAACAGTATATACA | 420       |              |
| Db                    | 361             | ATTATATCTTACATTTGTCAATATATAGGAAGTTTGTCCGTGGATGCAACAGTATATACA | 420       |              |

Db 361 ATTATATCTTACATTTGTCAAATATAAGGAAGTTTGTCCGTGGATGCAAAACAGTATATACA 420  
Qy 421 CAGGTGAGATCAAGCCAGACAGCTCTGGAAGTTCTTCTTACTTAATCTTAATCTTGGAAACA 480  
Db 421 CAAGTGAGATCAAGCCAGACAGCTCTGGAAGTTCTTCTTACTTAATCTTAATCTTGGAAACA 480  
Qy 481 ACATATGAATTAAGGTGTCTGCTGAAACACAGTGTGGCATTTGGAGTGTTTAGTGATCCA 540  
Db 481 ACATATGAATTAAGGTGTCTGCTGAAACACAGTGTGGCATTTGGAGTGTTTAGTGATCCA 540  
Qy 541 TTTCTCTTCCAACTGAGAAAGTGTCTGAGAAAGTGTGGAATCTCAAGTTGAGGCC 600  
Db 541 TTTCTCTTCCAACTGAGAAAGTGTCTGAGAAAGTGTGGAATCTCAAGTTGAGGCC 600  
Qy 601 TACAACGCTTCAGCAGTTAAGCTGATTTGGTATTTTACCTCGGCACCAATGGCAAAATTT 560  
Db 601 TACAACGCTTCAGCAGTTAAGCTGATTTGGTATTTTACCTCGGCACCAATGGCAAAATTT 560  
Qy 661 ACCAGCTTCAAGANTTAGTGTCAAACATGCCAGAAAGTGGATAGTAGTGAAGATGTCTCA 720  
Db 661 ACCAGCTTCAAGANTTAGTGTCAAACATGCCAGAAAGTGGATAGTAGTGAAGATGTCTCA 720  
Qy 721 ATCAGAGTAGAGACATTTTGAATCTGGGAAATTCGCCAGAAATGCCAGAAATAGTGAATCT 780  
Db 721 ATCAGAGTAGAGACATTTTGAATCTGGGAAATTCGCCAGAAATGCCAGAAATAGTGAATCT 780  
Qy 781 TTTTATGAGTAGACAGCCAGCCCTTCTCCAACTTGGTAGAGTTACACTCCATCGCGT 840  
Db 781 TTTTATGAGTAGACAGCCAGCCCTTCTCCAACTTGGTAGAGTTACACTCCATCGCGT 840  
Qy 841 ACCACACATTTCAAGCAGCTTGTGACACAGAAATGAGATCAGCTCTGTGTGGAAGAGCCT 900  
Db 841 ACCACACATTTCAAGCAGCTTGTGACACAGAAATGAGATCAGCTCTGTGTGGAAGAGCCT 900  
Qy 901 ATCAGTTTTGTAGTGACACATTTGAGACCTTATACAACATATCTTTTGAAGTTTCAGCT 960  
Db 901 ATCAGTTTTGTAGTGACACATTTGAGACCTTATACAACATATCTTTTGAAGTTTCAGCT 960  
Qy 961 GCTTACAACCTGAGCAGGTTATTTAGTAGTACGATTTGTGAGAACACAGAAATCAGTGCCT 1020  
Db 961 GCTTACAACCTGAGCAGGTTATTTAGTAGTACGATTTGTGAGAACACAGAAATCAGTGCCT 1020  
Qy 1021 GAAGGACCCACCAAACTGCGTAAACAGGCAACATCACAGGAAAGTCTTTTCAATTTTA 1080  
Db 1021 GAAGGACCCACCAAACTGCGTAAACAGGCAACATCACAGGAAAGTCTTTTCAATTTTA 1080  
Qy 1081 TGGGACCCACCAACTATAGTAAACAGGAAATTTAGTTATAGAGTTGAAATTAATGAGCCA 1140  
Db 1081 TGGGACCCACCAACTATAGTAAACAGGAAATTTAGTTATAGAGTTGAAATTAATGAGCCA 1140  
Qy 1141 TCAGGTGCGATTTTGGATTAACAGCACAAAGACCTCAAGTTTGCATTTCACTAACCTTAAACA 1200  
Db 1141 TCAGGTGCGATTTTGGATTAACAGCACAAAGACCTCAAGTTTGCATTTCACTAACCTTAAACA 1200  
Qy 1201 CCATTTTACAGTATGATGTCTATATTTGCGGTGAAACAGTGCAGGGACCTGGGCCCAAG 1260  
Db 1201 CCATTTTACAGTATGATGTCTATATTTGCGGTGAAACAGTGCAGGGACCTGGGCCCAAG 1260  
Qy 1261 TCAAAATTTTTCAGTATTTCACTCCACAGATGTTTCCAGGGGCGAGTGTGATTTTCAACTT 1320  
Db 1261 TCAAAATTTTTCAGTATTTCACTCCACAGATGTTTCCAGGGGCGAGTGTGATTTTCAACTT 1320  
Qy 1321 GCAGAGGTAGATTCACGCGAGTAAAGATTTTCTTGGAAAGAACCCAGCACCAATGGA 1380  
Db 1321 GCAGAGGTAGATTCACGCGAGTAAAGATTTTCTTGGAAAGAACCCAGCACCAATGGA 1380  
Qy 1381 ATTTATTTAAACCAATACCGAGTGAAGTGTAGTTTCCAGAGACAGGAATAATTTTGGAAAT 1440  
Db 1381 ATTTATTTAAACCAATACCGAGTGAAGTGTAGTTTCCAGAGACAGGAATAATTTTGGAAAT 1440  
Qy 1441 ACTTTGCTCACTGGAAATTAATGAGTATATAAATGACCCCTCCAGAAATTTGTGAAC 1500  
Db 1441 ACTTTGCTCACTGGAAATTAATGAGTATATAAATGACCCCTCCAGAAATTTGTGAAC 1500

Qy 1501 ATAGTAGAGCCAAATGTTAGGATTATATGAGGTTTCCAGCAGAGATGTGCTTCCACCTTCAC 1560  
Db 1501 ATAGTAGAGCCAAATGTTAGGATTATATGAGGTTTCCAGCAGAGATGTGCTTCCACCTTCAC 1560  
Qy 1561 TCATCTTGCTTACATTTATATATAACAGCCATCCAGATAAATACTTCTGCAAGGAATAGA 1620  
Db 1561 TCATCTTGCTTACATTTATATAACAGCCATCCAGATAAATACTTCTGCAAGGAATAGA 1620  
Qy 1621 GCTGAAGACCCAGACTTCCAGCTTGTAACTTACAGGAATCAGTATATTTACTGACATTGCA 1680  
Db 1621 GCTGAAGACCCAGACTTCCAGCTTGTAACTTACAGGAATCAGTATATTTACTGACATTGCA 1680  
Qy 1681 GCTGAACAGCTGTCTTATGTTTATCAGGAGACTTTGACCTTTCACCTGAGCACAATGATGT 1740  
Db 1681 GCTGAACAGCTGTCTTATGTTTATCAGGAGACTTTGACCTTTCACCTGAGCACAATGATGT 1740  
Qy 1741 GTATCTGCTTTTCCACCATCATGGGAGAGGACCAACACAGTTCCTCAGTGTTTAGGACAGT 1800  
Db 1741 GTATCTGCTTTTCCACCATCATGGGAGAGGACCAACACAGTTCCTCAGTGTTTAGGACAGT 1800  
Qy 1801 CAGCAAGTGCAGAGCTCCATTTAAATTTATAAACTATATAAAATATTAGTTCCTTCTATT 1860  
Db 1801 CAGCAAGTGCAGAGCTCCATTTAAATTTATAAACTATATAAAATATTAGTTCCTTCTATT 1860  
Qy 1861 TTTGTTATATTTGGGATCTCTCCAGATATCCCAATGAAAAATAACTCCTATACGATTTTAT 1920  
Db 1861 TTTGTTATATTTGGGATCTCTCCAGATATCCCAATGAAAAATAACTCCTATACGATTTTAT 1920  
Qy 1921 GCAATGGAATTTGGATATCAAAACAGAGCATTTCCAGATAAATACCATAGATAAAGCTTCTC 1980  
Db 1921 GCAATGGAATTTGGATATCAAAACAGAGCATTTCCAGATAAATACCATAGATAAAGCTTCTC 1980  
Qy 1981 ATAAACAGGGTTAAAGAAATACAAAAATACAAAAATGAGAGTGGAGCCCTCAACCCACGAT 2040  
Db 1981 ATAAACAGGGTTAAAGAAATACAAAAATACAAAAATGAGAGTGGAGCCCTCAACCCACGAT 2040  
Qy 2041 GGAGAAAGTTCTTCTGTGAGAAATGACATCTTTTGTGAGAACTTCAGAAAGTGAACCG 2100  
Db 2041 GGAGAAAGTTCTTCTGTGAGAAATGACATCTTTTGTGAGAACTTCAGAAAGTGAACCG 2100  
Qy 2101 GAATCATCACTTCAAGATGTGGAAGTAAATGATTTTACCGCAGATGAAATTAAGTTGAAG 2160  
Db 2101 GAATCATCACTTCAAGATGTGGAAGTAAATGATTTTACCGCAGATGAAATTAAGTTGAAG 2160  
Qy 2161 TGCTCACACCCGAAAGCCCAATGGGATCATTATTTGCTTATGAGTGTCTATATAAAT 2220  
Db 2161 TGCTCACACCCGAAAGCCCAATGGGATCATTATTTGCTTATGAGTGTCTATATAAAT 2220  
Qy 2221 ATAGATACCTTTTATATATGAAGAACACATCAACACAGACATATAATTAAGGAACCTTAAGA 2280  
Db 2221 ATAGATACCTTTTATATATGAAGAACACATCAACACAGACATATAATTAAGGAACCTTAAGA 2280  
Qy 2281 CCTCACACCTCTATAACATTTCTGTAAGGTCTTACACAGATTTTGGTCATGGCAATCAG 2340  
Db 2281 CCTCACACCTCTATAACATTTCTGTAAGGTCTTACACAGATTTTGGTCATGGCAATCAG 2340  
Qy 2341 GTATCTTCTTACTCTCTGTAAGGACTTCGGGAGACTGTGCGCTGTAGTGTGACACAGAAAT 2400  
Db 2341 GTATCTTCTTACTCTCTGTAAGGACTTCGGGAGACTGTGCGCTGTAGTGTGACACAGAAAT 2400  
Qy 2401 ATCACTTACAAAAATATTTCTCTCGAGAGATTGAGCTATCATTCTTCCCTCCCAAGTAGT 2460  
Db 2401 ATCACTTACAAAAATATTTCTCTCGAGAGATTGAGCTATCATTCTTCCCTCCCAAGTAGT 2460  
Qy 2461 CCCAATGGAATCATAAAAAATATACAAATTTATCTCAAGAGAAGTAAATGGAAATGAGGAA 2520  
Db 2461 CCCAATGGAATCATAAAAAATATACAAATTTATCTCAAGAGAAGTAAATGGAAATGAGGAA 2520  
Qy 2521 AGAAGTATTAATAACACCTTTTAAACCCAAACATTTAAAGTACTGAAAGAAATATATACCCAA 2580  
Db 2521 AGAAGTATTAATAACACCTTTTAAACCCAAACATTTAAAGTACTGAAAGAAATATATACCCAA 2580

QY 2581 TATATCAATTCAGGTGCTCTAGTACACTGAAAGGTGAGAGTTCGGAGTCTCCCAT 2640  
Db 2581 TATATCAATTCAGGTGCTCTAGTACACTGAAAGGTGAGAGTTCGGAGTCTCCCAT 2640  
QY 2641 AGTATACAGCGGAGGAAGTCTCTGATTCCTCCCTCAAGACTTCTCTGTAAGACAG 2700  
Db 2641 AGTATACAGCGGAGGAAGTCTCTGATTCCTCCCTCAAGACTTCTCTGTAAGACAG 2700  
QY 2701 TTGTCTGGTGTCAAGGTGTAAGTGTGATGCGAACACACCCCTGGAGCAAATGGAATATC 2760  
Db 2701 TTGTCTGGTGTCAAGGTGTAAGTGTGATGCGAACACACCCCTGGAGCAAATGGAATATC 2760  
QY 2761 CTTTATTACACAGTTTATGCTCGGAATAGATCATTAATAAATACTAATGTCAGTGA 2820  
Db 2761 CTTTATTACACAGTTTATGCTCGGAATAGATCATTAATAAATACTAATGTCAGTGA 2820  
QY 2821 ACATCATTTGAGTTATCAGATTTGGATTATAATGTTGAATACAGTCTTATGTAACAGCT 2880  
Db 2821 ACATCATTTGAGTTATCAGATTTGGATTATAATGTTGAATACAGTCTTATGTAACAGCT 2880  
QY 2881 AGCACAGATTTGGTGAATGGAAGAACAGGAAGCAATATCATTTAGCTTTCAACACAGAG 2940  
Db 2881 AGCACAGATTTGGTGAATGGAAGAACAGGAAGCAATATCATTTAGCTTTCAACACAGAG 2940  
QY 2941 GAGGACCAAGCGATCCTCCCAAGATGTTTATATGACCAACCTCAGTCTTTCATCAATA 3000  
Db 2941 GAGGACCAAGCGATCCTCCCAAGATGTTTATATGACCAACCTCAGTCTTTCATCAATA 3000  
QY 3001 ATTCTTTCTGGACACCTCTCTCAAAACCTTAATGGGATTATACAAATTTACTCTGTTAT 3060  
Db 3001 ATTCTTTCTGGACACCTCTCTCAAAACCTTAATGGGATTATACAAATTTACTCTGTTAT 3060  
QY 3061 TACAGAAATATCTTCAGGTACTTTTATGCGAAATTTTACACTCCATGAATCAATGAC 3120  
Db 3061 TACAGAAATATCTTCAGGTACTTTTATGCGAAATTTTACACTCCATGAATCAATGAC 3120  
QY 3121 TTGACAAATAGCTGTATCCCAATTTATAGATAAATGACAAATTTAGCTACTATACA 3180  
Db 3121 TTGACAAATAGCTGTATCCCAATTTATAGATAAATGACAAATTTAGCTACTATACA 3180  
QY 3181 TTTTGGTTAAACAGCAAGTACTTCAGTTGGAATGGAATGGAATGGAATGGAATGGAAT 3240  
Db 3181 TTTTGGTTAAACAGCAAGTACTTCAGTTGGAATGGAATGGAATGGAATGGAATGGAAT 3240  
QY 3241 GTATACACAGATCAAGACATACCTGAAAGGTTTGTGGAACCTGACTTTACGAATCCAT 3300  
Db 3241 GTATACACAGATCAAGACATACCTGAAAGGTTTGTGGAACCTGACTTTACGAATCCAT 3300  
QY 3301 TCGTCAACTGCAATAAATGTAAGTGGTCCCAACCGGCTCAACCAACCGTCTAGTCTTC 3360  
Db 3301 TCGTCAACTGCAATAAATGTAAGTGGTCCCAACCGGCTCAACCAACCGTCTAGTCTTC 3360  
QY 3361 TACTATGTTTCACTGATCTTACAGCAGACTCTCGGCATGTGAGCACCCTCTGTTACA 3420  
Db 3361 TACTATGTTTCACTGATCTTACAGCAGACTCTCGGCATGTGAGCACCCTCTGTTACA 3420  
QY 3421 TATGAGAGAGCATATTTTGTATAATCTGGAAAAATACATGATTATATATTAATAAT 3480  
Db 3421 TATGAGAGAGCATATTTTGTATAATCTGGAAAAATACATGATTATATATTAATAAT 3480  
QY 3481 ACTCCATCAACGAAAGGATTTCTCTGATACCTATACCTGCCCAGCTATACATCAAGCT 3540  
Db 3481 ACTCCATCAACGAAAGGATTTCTCTGATACCTATACCTGCCCAGCTATACATCAAGCT 3540  
QY 3541 GAAGAAGATGCCAGAACTTCAACCAATTAACACACTTTTAAACCTTTCTCTTACC 3600  
Db 3541 GAAGAAGATGCCAGAACTTCAACCAATTAACACACTTTTAAACCTTTCTCTTACC 3600  
QY 3601 TCAGTTCTCTTATCATGGATCCCCAGTAAGGCAAAATGTTGCAATATAAGTTATGAT 3660  
Db 3601 TCAGTTCTCTTATCATGGATCCCCAGTAAGGCAAAATGTTGCAATATAAGTTATGAT 3660  
QY 3661 TTAACCTTTACAGGACCAATGAAATTTATTTCTTCTTACTTCTGATAATTTACATAATA 3720

Db 3661 TTAACCTTTACAGGACCAATGAAATTTATTTCTTCTTACTTCTGATAATTTACATAATA 3720  
QY 3721 TTGGAAGAGCTTTTCAACATTTATATATAGCTTTTGTGCGCGCAAGAACTAGAAAA 3780  
Db 3721 TTGGAAGAGCTTTTCAACATTTATATATAGCTTTTGTGCGCGCAAGAACTAGAAAA 3780  
QY 3781 GGAATTTGCTTCCAGTATTTCTTTTCTTTTACAGATGAGTCACTGCGGTAGACCT 3840  
Db 3781 GGAATTTGCTTCCAGTATTTCTTTTCTTTTACAGATGAGTCACTGCGGTAGACCT 3840  
QY 3841 CCACAAAAATTTGACCTTTAATCACTGTTTCTCAGACTTTTGTATGGTGAATGGAGCCCA 3900  
Db 3841 CCACAAAAATTTGACCTTTAATCACTGTTTCTCAGACTTTTGTATGGTGAATGGAGCCCA 3900  
QY 3901 AGTCTCTTCCAGGTGGTATTGTTAAAGTATATAGTTTAAATTTTCAATGAAATGAACT 3960  
Db 3901 AGTCTCTTCCAGGTGGTATTGTTAAAGTATATAGTTTAAATTTTCAATGAAATGAACT 3960  
QY 3961 GACACTATATATTAAGAAATATATCAGGATTTAAATCTGAAGCCAAACTTGTGACTG 4020  
Db 3961 GACACTATATATTAAGAAATATATCAGGATTTAAATCTGAAGCCAAACTTGTGACTG 4020  
QY 4021 GAACAGTCAAGCCTTCTATCGTGTATCTGCTTCCAAAGTTGGAATGGCAAT 4080  
Db 4021 GAACAGTCAAGCCTTCTATCGTGTATCTGCTTCCAAAGTTGGAATGGCAAT 4080  
QY 4081 CAATTTAGTAATGTAAGTAAATTTCAACCCCAAGATCAGTTCCAGATGCTGTCAGAAAT 4140  
Db 4081 CAATTTAGTAATGTAAGTAAATTTCAACCCCAAGATCAGTTCCAGATGCTGTCAGAAAT 4140  
QY 4141 ATGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 4200  
Db 4141 ATGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 4200  
QY 4201 GCAATGGAATTAATTAACCGTATATGTTAGTGAAGAAATTTCTCAAAAGTTTCT 4260  
Db 4201 GCAATGGAATTAATTAACCGTATATGTTAGTGAAGAAATTTCTCAAAAGTTTCT 4260  
QY 4261 CCCCAAGATCAGTCACTTTTCAAAAGCTTTTGGCAATPACCTCATATGCTTTTAA 4320  
Db 4261 CCCCAAGATCAGTCACTTTTCAAAAGCTTTTGGCAATPACCTCATATGCTTTTAA 4320  
QY 4321 GTAAGAGCTTCAACCTCAGCTGGTGAAGTGAAGCACAATGCTGTCAGACACTA 4380  
Db 4321 GTAAGAGCTTCAACCTCAGCTGGTGAAGTGAAGCACAATGCTGTCAGACACTA 4380  
QY 4381 CCTGAAACAGTTCCTGAGTTCCTCAAAATATTTGCTTTTCTGATGTTCACTAGT 4440  
Db 4381 CCTGAAACAGTTCCTGAGTTCCTCAAAATATTTGCTTTTCTGATGTTCACTAGT 4440  
QY 4441 GCAACATTTGACATGATGAAGTCACTGACACTTCTTGGCTACTTTTCAAAATTTACAAAT 4500  
Db 4441 GCAACATTTGACATGATGAAGTCACTGACACTTCTTGGCTACTTTTCAAAATTTACAAAT 4500  
QY 4501 ACCACTCAACTTCTGCTCAAAATGCAAGTGGAAATGCAAGAAATGCTGTAATAT 4560  
Db 4501 ACCACTCAACTTCTGCTCAAAATGCAAGTGGAAATGCAAGAAATGCTGTAATAT 4560  
QY 4561 CAAAAATTTCAATACCTCTATGAAGTCACTTAACTGAAGAGACAGTATATGGATTAAG 4620  
Db 4561 CAAAAATTTCAATACCTCTATGAAGTCACTTAACTGAAGAGACAGTATATGGATTAAG 4620  
QY 4621 AAATTTAGTATGATAGATTTCCAGTGGTCCAGCAATGCTGGCTATGGCAATGCT 4680  
Db 4621 AAATTTAGTATGATAGATTTCCAGTGGTCCAGCAATGCTGGCTATGGCAATGCT 4680  
QY 4681 TCAAACTGGATTTCTACAAAAACTCTGCTGGCTCCAGATGGTCTCTCTGAAATGTT 4740  
Db 4681 TCAAACTGGATTTCTACAAAAACTCTGCTGGCTCCAGATGGTCTCTCTGAAATGTT 4740  
QY 4741 CATGTAGTAGCAACATCACTTTTAGCATCAGCATAGCTGGAGTGAACCTGCTGCTCAT 4800

4741 CATGTAGTAGCAACATCACTCTTTTAGCATCAGCATAAAGCTGGAGTGAACCTGCTGTCTATT 4800  
4801 ACTGGACCAACATGTTTCTGATTTGATGTCAAATCGGTAGATAATGATGAATTTAATATA 4860  
4801 ACTGGACCAACATGTTTCTGATTTGATGTCAAATCGGTAGATAATGATGAATTTAATATA 4860  
4861 TCCTTCATCAAGTCAAAATGAAGAAAAATAAACCATAGAAAATTAAGAAATTTATTC 4920  
4861 TCCTTCATCAAGTCAAAATGAAGAAAAATAAACCATAGAAAATTAAGAAATTTATTC 4920  
4921 ACAAGTATTCGTAGTATGATCTGATTTACTGCGAAATTTAGGAAATTTAGGAAATTTAG 4980  
4921 ACAAGTATTCGTAGTATGATCTGATTTACTGCGAAATTTAGGAAATTTAGGAAATTTAG 4980  
4981 GGGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 5040  
4981 GGGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 5040  
5041 AACCAATGACATTTTCAGAAAGATACCAGATGAAGTTTACAAAATTTCAATTAACGTTCCCT 5100  
5041 AACCAATGACATTTTCAGAAAGATACCAGATGAAGTTTACAAAATTTCAATTAACGTTCCCT 5100  
5101 CCTCCTTCTCAACCTAATGGAATATCCAAAGTATATCAAGTCTGCTTTTACCGAGAGAT 5160  
5101 CCTCCTTCTCAACCTAATGGAATATCCAAAGTATATCAAGTCTGCTTTTACCGAGAGAT 5160  
5161 GATCCTACTGCTGCTCAGATTCACAACTCAGTATTTATACAGAAACCAACCACTTCGTC 5220  
5161 GATCCTACTGCTGCTCAGATTCACAACTCAGTATTTATACAGAAACCAACCACTTCGTC 5220  
5221 ATTGCAATGCTAGAAAGCTAAAAGGTGGACATACATCAATATCAAGTCTTTACGCGAGTC 5280  
5221 ATTGCAATGCTAGAAAGCTAAAAGGTGGACATACATCAATATCAAGTCTTTACGCGAGTC 5280  
5281 AATAGTCTGGTGCAGGTCCAAAGTTCGATGAGAAATTAACCATGATATCAAAAGTCCCA 5340  
5281 AATAGTCTGGTGCAGGTCCAAAGTTCGATGAGAAATTAACCATGATATCAAAAGTCCCA 5340  
5341 GCAGACCAAAACCAACCCCTATTTATGATGCCACAGGAAATCTGTTGTGACT 5400  
5341 GCAGACCAAAACCAACCCCTATTTATGATGCCACAGGAAATCTGTTGTGACT 5400  
5401 TCAACCAACATTAACAATCAGAATGCCAATGTTACTACAGTATGATCATGAGCCCAATA 5460  
5401 TCAACCAACATTAACAATCAGAATGCCAATGTTACTACAGTATGATCATGAGCCCAATA 5460  
5461 AAAAAATGTAAGTGTGCGACAGAAACAGGAGCTCAGCATGATGGAATGTAAACAAAG 5520  
5461 AAAAAATGTAAGTGTGCGACAGAAACAGGAGCTCAGCATGATGGAATGTAAACAAAG 5520  
5521 TGGTATGATGCATATTTTAAATGAAGCAAGCCCATATTTTACAAATGAAGCTTTCCCTAAC 5580  
5521 TGGTATGATGCATATTTTAAATGAAGCAAGCCCATATTTTACAAATGAAGCTTTCCCTAAC 5580  
5581 CCTCCATGTACAGAGGAAGACAAAAGTTTAGTGGCAATGAAGAAATCTACATCATAGGT 5640  
5581 CCTCCATGTACAGAGGAAGACAAAAGTTTAGTGGCAATGAAGAAATCTACATCATAGGT 5640  
5641 GCTGATAATGATGCATGATTTCTCGCAATGAAGACAAAATTTGCAATGGAACCACTGAAA 5700  
5641 GCTGATAATGATGCATGATTTCTCGCAATGAAGACAAAATTTGCAATGGAACCACTGAAA 5700  
5701 CCAAAAAGCAATACCTATTTTAAATTTAGAGCTACAAATATTTATGGGCAATTTACTGAC 5760  
5701 CCAAAAAGCAATACCTATTTTAAATTTAGAGCTACAAATATTTATGGGCAATTTACTGAC 5760  
5761 TCTGATATTTCTGACCTGTTTAAAGCTTTTAGGGGAAGGACTTTCAGAAAGAACCGTAGAG 5820  
5761 TCTGATATTTCTGACCTGTTTAAAGCTTTTAGGGGAAGGACTTTCAGAAAGAACCGTAGAG 5820  
5821 ATCAATCTTTCCGTCACTTTGTGATTCCTTTCAATAATTTCTCCTGGAAACAGCTATTTT 5880  
5821 ATCAATCTTTCCGTCACTTTGTGATTCCTTTCAATAATTTCTCCTGGAAACAGCTATTTT 5880

5881 GCATTTGCAAGAAATTCGACAGAACGAAAGGTGGCACATATCTCTCTCAGGATGCA 5940  
5881 GCATTTGCAAGAAATTCGACAGAACGAAAGGTGGCACATATCTCTCTCAGGATGCA 5940  
5941 GAAATTAATGACATTAATTAAGCTGATCAGCTCATCAAGTGGGAGACCTGGAATG 6000  
5941 GAAATTAATGACATTAATTAAGCTGATCAGCTCATCAAGTGGGAGACCTGGAATG 6000  
6001 AAGGACGAGAGNTTAAACGGGCCAATAGCAGAAATCTCTCTGCAACATGTTGAAGAG 6060  
6001 AAGGACGAGAGNTTAAACGGGCCAATAGCAGAAATCTCTCTGCAACATGTTGAAGAG 6060  
6061 CTTTTCACAAAACAACTAAAGTTTCAAGAAAGAAATTTTGGAAATTAACCAAAATTTCTT 6120  
6061 CTTTTCACAAAACAACTAAAGTTTCAAGAAAGAAATTTTGGAAATTAACCAAAATTTCTT 6120  
6121 CAGGATCTTTTCAACTGATGCTGATCTGCTTGAATAGAGCAAAAACCGTTTCCCA 6180  
6121 CAGGATCTTTTCAACTGATGCTGATCTGCTTGAATAGAGCAAAAACCGTTTCCCA 6180  
6181 AACATAAACCAATATAATAATAAACAAGCTAAAGCTGATAGCTGACGCTAGTGTCCA 6240  
6181 AACATAAACCAATATAATAATAAACAAGCTAAAGCTGATAGCTGACGCTAGTGTCCA 6240  
6241 GGTTCGGATTTATTTAAATGCCAGCTATATTTCTGCTTATTTATGTCCTCAAAATGAAATTT 6300  
6241 GGTTCGGATTTATTTAAATGCCAGCTATATTTCTGCTTATTTATGTCCTCAAAATGAAATTT 6300  
6301 GCTACTCAAGGTCCACTACAGGAAACAGTTGGAGATTTTGGAGAAATGGTGTGGGAAACC 6360  
6301 GCTACTCAAGGTCCACTACAGGAAACAGTTGGAGATTTTGGAGAAATGGTGTGGGAAACC 6360  
6361 AGGGCAAAAACATTTAGTAAATGCTAAACAGTGTGTTTGAAGAGGAGCTGATGAGTCCAT 6420  
6361 AGGGCAAAAACATTTAGTAAATGCTAAACAGTGTGTTTGAAGAGGAGCTGATGAGTCCAT 6420  
6421 CAGTATTTGGCCAGAGACAAACAGGACGTTACTGCTTTGGAGATATAGTGAATTAACAAG 6480  
6421 CAGTATTTGGCCAGAGACAAACAGGACGTTACTGCTTTGGAGATATAGTGAATTAACAAG 6480  
6481 CTAATGGAGGATGTTCAAATAGTTGACTATCAGGATCTGAAAATTTGAAAGGATGGG 6540  
6481 CTAATGGAGGATGTTCAAATAGTTGACTATCAGGATCTGAAAATTTGAAAGGATGGG 6540  
6541 GATTCATGACGTTTTCGACAGTGTAACTTTACTGCTGCTGGCCAGAGCATGGGTTCTGAG 6600  
6541 GATTCATGACGTTTTCGACAGTGTAACTTTACTGCTGCTGGCCAGAGCATGGGTTCTGAG 6600  
6601 AACAGCGCCCTCTAAATTCACCTTTCTGAAAGTTGGTTTCGAGCAAGCAGGCAATGACACC 6660  
6601 AACAGCGCCCTCTAAATTCACCTTTCTGAAAGTTGGTTTCGAGCAAGCAGGCAATGACACC 6660  
6661 ACACCTATGATGTTTCACTGCTGAGTGTGAGTGAAGAACTGAGGATTTTATGCTCTG 6720  
6661 ACACCTATGATGTTTCACTGCTGAGTGTGAGTGAAGAACTGAGGATTTTATGCTCTG 6720  
6721 GACCATTTAACACACATATAAATGACCATGATTTTGGATATATATGACCTAGTGTG 6780  
6721 GACCATTTAACACACATATAAATGACCATGATTTTGGATATATATGACCTAGTGTG 6780  
6781 GAACTGAGAGGTGAAGAAATGTCATGCTGAGAAATCTGGCACAGTATATCTTTTACAC 6840  
6781 GAACTGAGAGGTGAAGAAATGTCATGCTGAGAAATCTGGCACAGTATATCTTTTACAC 6840  
6841 CAGTGCATTTCTGATCTCTTATCAATTAAGGGAAGTAATCAGCCCATCTGTTTGTAAAC 6900  
6841 CAGTGCATTTCTGATCTCTTATCAATTAAGGGAAGTAATCAGCCCATCTGTTTGTAAAC 6900  
6901 TATTCAGCACTTCAGAGNTGCACTCTTTGGAGCCCATGGAAGGTGATGTTGAGCTTGAA 6960  
6901 TATTCAGCACTTCAGAGNTGCACTCTTTGGAGCCCATGGAAGGTGATGTTGAGCTTGAA 6960

QY 6961 TGGGAAGAACCACTATGTAATAATTCAGACCAAGGATACAAATGGAAGAGATTTTAA 7020  
Db |||||  
6961 TGGGAAGAACCACTATGTAATAATTCAGACCAAGGATACAAATGGAAGAGATTTTAA 7020  
QY 7021 ATCCAGGGGCCAAAGTTACCCCTCATTTCTCCGAATTCGAATGTCGAACCTTAAAGAA 7080  
Db |||||  
7021 ATCCAGGGGCCAAAGTTACCCCTCATTTCTCCGAATTCGAATGTCGAACCTTAAAGAA 7080  
QY 7081 ATATCTATGCTTCTCTCACTGTGCTTT 7108  
Db |||||  
7081 ATATCTATGCTTCTCTCACTGTGCTTT 7108

## RESULT 2

US-08-036-210-21  
; Sequence 21, Application US/08036210  
; Patent No. 5585233  
; GENERAL INFORMATION:  
; APPLICANT: Moller, Niels P.H.  
; APPLICANT: Moller, Karin B.  
; APPLICANT: Ullrich, Axel  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASE  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/036,210  
; FILING DATE: 23-MAR-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mierock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7683-025  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3973 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 3430\_  
; OTHER INFORMATION: /note= "N=x-unknown nucleotide"  
US-08-036-210-21

Query Match 36.8%; Score 2618.6; DB 1; Length 3973;  
Best Local Similarity 97.2%; Pred. No. 0;  
Matches 2716; Conservative 0; Mismatches 4; Indels 75; Gaps 2;  
QY 4386 AACAGTCCAGGTGTCACCAAAATATGCTTTTCTGATGTTCACTCACTAGTGCAC 4445  
Db 33 AAAAGTCCAGGTGTCACCAAAATATGCTTTTCTGATGTTCACTCACTAGTGCAC 92  
QY 4446 ATTGACATGATGAAGACCTGACACTATCTTGCTACTTTCAAATTTACAAATTTACCAC 4505  
Db |||||  
93 ATTGACATGATGAAGACCTGACACTATCTTGCTACTTTCAAATTTACAAATTTACCAC 152

QY 4506 TCACTTCTGCTCAAAAATGCAAGATCGGAATCCGAAGATGTGTTGAATATCAAAA 4565  
Db |||||  
153 TCACTTCTGCTCAAAAATGCAAGATCGGAATCCGAAGATGTGTTGAATATCAAAA 212  
QY 4566 AATTCAATACCTCTATGAAGCTCAGTTAACTGAAGAGACAGATATATGGATTAAGAAAT 4625  
Db |||||  
213 AATTCAATACCTCTATGAAGCTCAGTTAACTGAAGAGACAGATATATGGATTAAGAAAT 272  
QY 4626 TAGATGGTATAGATTCGAAGTGGCTGCCAGACCAATGCTGGCTATGGCAATGCTTCAAA 4685  
Db |||||  
273 TAGATGGTATAGATTCGAAGTGGCTGCCAGACCAATGCTGGCTATGGCAATGCTTCAAA 332  
QY 4686 CTGGATTTCTACAAAACCTCTGCTGGCCCTCCAGATGCTCTCTCGAAAATGTTCAATGT 4745  
Db |||||  
333 CTGGATTTCTACAAAACCTCTGCTGGCCCTCCAGATGCTCTCTCGAAAATGTTCAATGT 392  
QY 4746 AGTAGCAACATCACCTTTTAGCATCAGCATAGCTGGAGTGAACCTGCTGCTCAATCTGG 4805  
Db |||||  
393 AGTAGCAACATCACCTTTTAGCATCAGCATAGCTGGAGTGAACCTGCTGCTCAATCTGG 452  
QY 4806 ACCAACAATGTTATCTGATGATGTCAAATCGGTAGATAATGATGAATTTAATAATATCTCT 4865  
Db |||||  
453 ACCAACAATGTTATCTGATGATGTCAAATCGGTAGATAATGATGAATTTAATAATATCTCT 512  
QY 4866 CATCAAGTCAAAATGAAGAAAATAAACCATAGAAAATTAAGATTTAGAAAATATTCACAG 4925  
Db |||||  
513 CATCAAGTCAAAATGAAGAAAATAAACCATAGAAAATTAAGATTTAGAAAATATTCACAG 572  
QY 4926 GTATTCTGTAGTATCAGCTGCAATTTACTGGGAACATTAGTGTGCTGATATGTAGAGGAA 4985  
Db |||||  
573 GTATTCTGTAGTATCAGCTGCAATTTACTGGGAACATTAGTGTGCTGATATGTAGAGGAA 632  
QY 4986 GTCAAGTGTGAAATGATTTGTTACTATTAGAAATCAGCCCAAGAGGCCACCTAACAA 5045  
Db |||||  
633 GTCAAGTGTGAAATGATTTGTTACTATTAGAAATCAGCCCAAGAGGCCACCTAACAA 692  
QY 5046 CATGACATTTAGAGATACCAAGTGAAGTTACAAAATTTCAATTAAGCTTCTCTCTCC 5105  
Db |||||  
693 CATGACATTTAGAGATACCAAGTGAAGTTACAAAATTTCAATTAAGCTTCTCTCTCC 752  
QY 5106 TTCTCAACCTAATGGAATATCAAGTATATCAAGCTCTGTTTACCGAGAGAGATGATCC 5165  
Db |||||  
753 TTCTCAACCTAATGGAATATCAAGTATATCAAGCTCTGTTTACCGAGAGAGATGATCC 812  
QY 5166 TACTGCTGTCAGATTCACAACTCAGTATTTATACAGAAAACCAACACATTCGTCATTGC 5225  
Db |||||  
813 TACTGCTGTCAGATTCACAACTCAGTATTTATACAGAAAACCAACACATTCGTCATTGC 872  
QY 5226 AATGCTAGAGGACTAAAGGTGGACATACATAATATCAGTGTGTTACGAGTCAATAG 5285  
Db |||||  
873 AATGCTAGAGGACTAAAGGTGGACATACATAATATCAGTGTGTTACGAGTCAATAG 932  
QY 5286 TGCTGTCGAGTCCAAAGGTTCCGATGAGATACCAATGATATCAAAAGTCCAGCAG 5345  
Db |||||  
933 TGCTGTCGAGTCCAAAGGTTCCGATGAGATACCAATGATATCAAAAGTCCAGCAG 992  
QY 5346 ACCAAAAACCAACCAACCCCTATTATATGATGCCACAGGAAAATGCTTGTGACTTCAAC 5405  
Db |||||  
993 ACCAAAAACCAACCAACCCCTATTATATGATGCCACAGGAAAATGCTTGTGACTTCAAC 1052  
QY 5406 AACAAATTAATCAGATGCCAAATATGTTTACAGTATGATGATGGAACCAATAAAAA 5465  
Db |||||  
1053 AACAAATTAATCAGATGCCAAATATGTTTACAGTATGATGATGGAACCAATAAAAA 1112  
QY 5466 TGTAAGTGTGTCGACAGAAACAGGAGCTCAGCATGATGATGGAATGTAACAAAGTGA 5525  
Db |||||  
1113 TGTAAGTGTGTCGACAGAAACAGGAGCTCAGCATGATGATGGAATGTAACAAAGTGA 1172  
QY 5526 TGATGATATTTTAAATGAAGGCCATATTTTAAATGAAGGCTTTCTTAACCTCC 5585  
Db |||||  
1173 TGATGATATTTTAAATGAAGGCCATATTTTAAATGAAGGCTTTCTTAACCTCC 1232  
QY 5586 ATGTAAGAGGAAAGACAAAGTGTGTCGAATGAAGAAATCTACATCATAGGTGCTGA 5645

Db 1233 ATGTACAGAGGAAACACAAAGTTTAGTGGCAATGAAGAAATCTACATCATAGGTGCTGA 1292  
Qy 5646 TAATGATGATGATTTCTGGCAATGAAGACAAAATTTGCAATGGACCACTGAACCAAA 5705  
Db 1293 TAATGATGATGATTTCTGGCAATGAAGACAAAATTTGCAATGGACCACTGAACCAAA 1352  
Qy 5706 AAAGCAATCTTATTTAAATTTAGAGCTACAAATATTTAGGACAAATTTACTGACTCTGA 5765  
Db 1353 AAAGCAATCTTATTTAAATTTAGAGCTACAAATATTTAGGACAAATTTACTGACTCTGA 1412  
Qy 5766 TTATTTCTGACCCCTGTTAAGACTTTAGGGAAGGACTTTTCAGAAAGAACCGTAGAGATCAT 5825  
Db 1413 TTATTTCTGACCCCTGTTAAGACTTTAGGGAAGGACTTTTCAGAAAGAACCGTAGAGATCAT 1472  
Qy 5826 TCTTTCCGTCACCTTTGTGTATCTCTTCAATAATCTCTCTTGGAAAGAGCTATTTTGTGCTAT 5885  
Db 1473 TCTTTCCGTCACCTTTGTGTATCTCTTCAATAATCTCTCTTGGAAAGAGCTATTTTGTGCTAT 1532  
Qy 5886 TGCAAGAAATTCGACAGAGCAAGAAAGAGTGSCACATCTCTCTCAGGATGCGAAT 5945  
Db 1533 TGCAAGAAATTCGACAGAGCAAGAAAGAGTGSCACATCTCTCTCAGGATGCGAAT 1592  
Qy 5946 TATTTGACACTAAATTTGAAGCTGGATCAGCTCATCAAGTGGCAGACCTTGGAACTGAAGGA 6005  
Db 1593 TATTTGACACTAAATTTGAAGCTGGATCAGCTCATCAAGTGGCAGACCTTGGAACTGAAGGA 1652  
Qy 6006 CGAGAGATTAACGCG----- 6020  
Db 1653 CGAGAGATTAACGCGTACTCTTTCAATTTTCTTAGACGCAAGGAGATTTTGTCTATCCA 1712  
Qy 6021 -----GCCAATAAGCAAGAAATCTCTCTCAGCAATCT 6053  
Db 1713 GTTACTTAGTTATAGAAAATCCATCAAGCCAAATAGCAAGAAATCTCTCTGCAATGT 1772  
Qy 6054 TGAAGAGCTTTGACAAAACAAACCTAAAGTTTCAAGAAAGAAATTTTCGGAATTAACAAA 6113  
Db 1773 TGAAGAGCTTTGACAAAACAAACCTAAAGTTTCAAGAAAGAAATTTTCGGAATTAACAAA 1832  
Qy 6114 ATTTCTTCAGGATCTTTCTTCAACTGATGCTGATCTGCTTGGAAATAGAGCAAAAACCG 6173  
Db 1833 ATTTCTTCAGGATCTTTCTTCAACTGATGCTGATCTGCTTGGAAATAGAGCAAAAACCG 1892  
Qy 6174 TTTCCCAACATAAACCATATAATAATAACAGATAAAGCTGATAGCTGACGCTAG 6233  
Db 1893 CTTCCCAACATAAACCAT- --ATAATAACAGATAAAGCTGATAGCTGACGCTAG 1949  
Qy 6234 TGTTCCAGGTTCCGATTTATTAATGCCAGCTATATTTCTGGTTATTTATGTCCAAATGA 6293  
Db 1950 TGTTCCAGGTTCCGATTTATTAATGCCAGCTATATTTCTGGTTATTTATGTCCAAATGA 2009  
Qy 6294 ATTTATTTGCTATCAAGGTTCCATACAGGAAACAGTTGGAGATTTTGGAGAAATGTGTG 6353  
Db 2010 ATTTATTTGCTATCAAGGTTCCATACAGGAAACAGTTGGAGATTTTGGAGAAATGTGTG 2069  
Qy 6354 GGAACACGCGGCAAAACATTAAGTATGCTTAACACAGTCTTTTGAAGAGCGGATCAG 6413  
Db 2070 GGAACACGAGCAAAACATTAAGTATGCTTAACACAGTCTTTTGAAGAGCGGATCAG 2129  
Qy 6414 ATGCCATCAGTATTTGGCCAGAGCAACAGGCAAGTCTGCTTTTGGAGATATAGTAT 6473  
Db 2130 ATGCCATCAGTATTTGGCCAGAGCAACAGGCAAGTCTGCTTTTGGAGATATAGTAT 2189  
Qy 6474 TACAAAGCTAATGGAGATGTTCAATAGATTGGAATATCAGGATCTGAAATTTGAAAG 6533  
Db 2190 TACAAAGCTAATGGAGATGTTCAATAGATTGGAATATCAGGATCTGAAATTTGAAAG 2249  
Qy 6534 GCATGGGATGATGCTGCTTCAAGTGTAACTTTTACTGCTGCGCAGAGCATGGGT 6593  
Db 2250 GCATGGGATGATGCTGCTTCAAGTGTAACTTTTACTGCTGCGCAGAGCATGGGT 2309  
Qy 6594 TCTGTGAGAACAGCGCCCTCTAAATTCATTTGTGAAGTTGGTTTCGAGCAAGCAGGCGACA 6653

Db 2310 TCTGTGAGAACAGCGCCCTCTAAATTCATTTGTGTGAAGTTGGTTTCGAGCAAGCAGGCGACA 2369  
Qy 6654 TGACACCAACACCTATGATTTGTTCACTGCAAGTCTGAGTGTGGAAGAACTGGAGTTTTTAT 6713  
Db 2370 TGACACCAACACCTATGATTTGTTCACTGCAAGTCTGAGTGTGGAAGAACTGGAGTTTTTAT 2429  
Qy 6714 TCTCTGTGACCACTTTAAACACCAACATATAAATGACCATGATTTTGTGGATATATATGGACT 6773  
Db 2430 TCTCTGTGACCACTTTAAACACCAACATATAAATGACCATGATTTTGTGGATATATATGGACT 2489  
Qy 6774 AGTAGCTGAACTGAGAGTGAAGAAATGTGCAATGTGCAAGTCTGGAATCTGGCAAGTATATCTT 6833  
Db 2490 AGTAGCTGAACTGAGAGTGAAGAAATGTGCAATGTGCAAGTCTGGCAAGTATATCTT 2549  
Qy 6834 TTTTACACAGTGCATTTCTGGATCTCTTATCAATAAGGGAAGTAAATCAGCCCATCTGTTT 6893  
Db 2550 TTTTACACAGTGCATTTCTGGATCTCTTATCAATAAGGGAAGTAAATCAGCCCATCTGTTT 2609  
Qy 6894 TGTTAACATATTTCAGCACTTTTCAGAGATGGACTCTTTTGGACGCCATGGAAGTGTATGTTGA 6953  
Db 2610 TGTTAACATATTTCAGCACTTTTCAGAGATGGACTCTTTTGGACGCCATGGAAGTGTATGTTGA 2669  
Qy 6954 GCTTGAATGGGAAGAAACCACTATGTAAATATTCAGACCAAAAGGATACAAATTTGGAAGAGA 7013  
Db 2670 GCTTGAATGGGAAGAAACCACTATGTAAATATTCAGACCAAAAGGATACAAATTTGGAAGAGA 2729  
Qy 7014 TTTTAAATCCCAGGGGCAAGTTTACCCCTCATTTCCGAATTTGAAATGTCGAACCT 7073  
Db 2730 TTTTAAATCCCAGGGGCAAGTTTACCCCTCATTTCCGAATTTGAAATGTCGAACCT 2789  
Qy 7074 TAAAGAAATATCTATGCTTTCTCTCACTGTGCTTTT 7108  
Db 2790 TAAAGAAATATCTATGCTTTCTCTCACTGTGCTTTT 2824

## RESULT 3

US-08-449-609-21  
; Sequence 21, Application US/08449609  
; Patent No. 5952212  
; GENERAL INFORMATION:  
; APPLICANT: Moller, Niels P.H.  
; APPLICANT: Moller, Karin B.  
; APPLICANT: Ullrich, Axel  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASE  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/449,609  
; FILING DATE: 24-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/036,210  
; FILING DATE: 23-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7683-025  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864/9741  
; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3973 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 3430  
; OTHER INFORMATION: /note= "N=x=unknown nucleotide"  
; US-08-449-609-21

Query Match 36.8%; Score 2618.6; DB 2; Length 3973;  
Best Local Similarity 97.2%; Pred. No. 0;  
Matches 2716; Conservative 0; Mismatches 4; Indels 75; Gaps 2;  
  
QY 4386 AACAGTTCCTCCAGGTTCCCAAAATATGCTTTTCTGATGTTTCAGTCACTAGTGCAC 4445  
DB 33 AAAAGTTCCTCCAGGTTCCCAAAATATGCTTTTCTGATGTTTCAGTCACTAGTGCAC 92  
  
QY 4446 ATTGACATGGATGAAGACCTGACACTATCTCTGGCTACTTTTCAAAATTTACAAAATTACCAC 4505  
DB 93 ATTGACATGGATGAAGACCTGACACTATCTCTGGCTACTTTTCAAAATTTACAAAATTACCAC 152  
  
QY 4506 TCAACTTCGTGCTCAAAAATGCAAGATGGGAATCGGAATCGGAATGTGTTGAATATCAAAA 4565  
DB 153 TCAACTTCGTGCTCAAAAATGCAAGATGGGAATCGGAATCGGAATGTGTTGAATATCAAAA 212  
  
QY 4566 AATTCATACCTCTATGAAGCTCAGTTTAACTGAAGACAGATATATGGAATTAAGAATTT 4625  
DB 213 AATTCATACCTCTATGAAGCTCAGTTTAACTGAAGACAGATATATGGAATTAAGAATTT 272  
  
QY 4626 TAGATGGTATAGATTCCTCAAGTGGCTGCCAGCAATGCTGGCTATGGCAATGCTTCAAA 4685  
DB 273 TAGATGGTATAGATTCCTCAAGTGGCTGCCAGCAATGCTGGCTATGGCAATGCTTCAAA 332  
  
QY 4686 CTGGATTTCTCAAAAATCTGCTGCTGCCCTCCAGATGCTCTCTGGAATGTTTCATGT 4745  
DB 333 CTGGATTTCTCAAAAATCTGCTGCTGCCCTCCAGATGCTCTCTGGAATGTTTCATGT 392  
  
QY 4746 AGTAGCAATCACTTTAGCATCAGCATAGCTGGAGTGAACCTGCTCATTAATCTGG 4805  
DB 393 AGTAGCAATCACTTTAGCATCAGCATAGCTGGAGTGAACCTGCTCATTAATCTGG 452  
  
QY 4806 ACCAATGTTATCTGATGATGTCATGTCGATGATGATGATGATGATGATGATGATGAT 4865  
DB 453 ACCAATGTTATCTGATGATGTCATGTCGATGATGATGATGATGATGATGATGATGAT 512  
  
QY 4866 CATCAAGTCAAAATGAAGAAATAAACCATAGAAATTTAAGATTTAAGAAATTTCAAAAG 4925  
DB 513 CATCAAGTCAAAATGAAGAAATAAACCATAGAAATTTAAGATTTAAGAAATTTCAAAAG 572  
  
QY 4926 GTATTCGTAGTATCACTGCTATTTACTGGACATTTAGTCTGCTGATATGTAGAGGGAA 4985  
DB 573 GTATTCGTAGTATCACTGCTATTTACTGGACATTTAGTCTGCTGATATGTAGAGGGAA 632  
  
QY 4986 GTCAAGTGTGAAATGATTTACTTACTTTAGAAATCAGCCCAAGGACCCACTTAACAA 5045  
DB 633 GTCAAGTGTGAAATGATTTACTTACTTTAGAAATCAGCCCAAGGACCCACTTAACAA 692  
  
QY 5046 CATGACATTTCAAGATACCAAGATGAAGTTACAAAATTTCAATTAACGTTCTCTCTCC 5105  
DB 693 CATGACATTTCAAGATACCAAGATGAAGTTACAAAATTTCAATTAACGTTCTCTCTCC 752  
  
QY 5106 TTCTCAACCTTAATGGAATATCAAGATATCAAGCTCTGGTTTACCGAAGATGATCC 5165  
DB 753 TTCTCAACCTTAATGGAATATCAAGATATCAAGCTCTGGTTTACCGAAGATGATCC 812  
  
QY 5166 TACTGTGCTCCAGATTCACAACTCAGTATATACAGAAACCAACACATTCGTCATTGC 5225  
DB 813 TACTGTGCTCCAGATTCACAACTCAGTATATACAGAAACCAACACATTCGTCATTGC 872

QY 5226 AATGCTAGAGGACTAAAGGTGGACATACATACATATCAGTGTTTACGCACTCAATAG 5285  
DB 873 AATGCTAGAGGACTAAAGGTGGACATACATACATATCAGTGTTTACGCACTCAATAG 932  
  
QY 5286 TGCTGTGTCAGGTCCAAAGGTTCCGATGAGATTAACCATGATATCAAGCTCCAGCAGC 5345  
DB 933 TGCTGTGTCAGGTCCAAAGGTTCCGATGAGATTAACCATGATATCAAGCTCCAGCAGC 992  
  
QY 5346 ACCAAAAACCAACCAACCCCTATTTATGATGCCACAGGAAAACTCTCTGTGTGCTTCAAC 5405  
DB 993 ACCAAAAACCAACCAACCCCTATTTATGATGCCACAGGAAAACTCTCTGTGTGCTTCAAC 1052  
  
QY 5406 AACCAATTACCAATGCAATGCAATGCTTACTACAGTGTGATCATGCGCAATTAACAAA 5465  
DB 1053 AACCAATTACCAATGCAATGCAATGCTTACTACAGTGTGATCATGCGCAATTAACAAA 1112  
  
QY 5466 TGTACAAAGTCTTGGCAGCAAGAACAGAGCTCAGCATGATGGAATGTAACAAAGTGGTA 5525  
DB 1113 TGTACAAAGTCTTGGCAGCAAGAACAGAGCTCAGCATGATGGAATGTAACAAAGTGGTA 1172  
  
QY 5526 TGATGCAATTTTAAATAAGCAAGGCCATATTTTACAAATGAAGGCTTTCTTAACCCCTCC 5585  
DB 1173 TGATGCAATTTTAAATAAGCAAGGCCATATTTTACAAATGAAGGCTTTCTTAACCCCTCC 1232  
  
QY 5586 ATGTACAGAGGAAAGACAAAGTTTAGTGGCAATGAAGAAATCTACATCATAGTGTCTGA 5645  
DB 1233 ATGTACAGAGGAAAGACAAAGTTTAGTGGCAATGAAGAAATCTACATCATAGTGTCTGA 1292  
  
QY 5646 TAATGATCATCATGATTTCTTGGCAATGAAGCAAAATTTGCAATGGACCACTGAAGAACCAAA 5705  
DB 1293 TAATGATCATCATGATTTCTTGGCAATGAAGCAAAATTTGCAATGGACCACTGAAGAACCAAA 1352  
  
QY 5706 AAAGCAATCTTATTTAAATTTAGAGCTACAAATATTTAGGGAATTTTACTGACTCTGA 5765  
DB 1353 AAAGCAATCTTATTTAAATTTAGAGCTACAAATATTTAGGGAATTTTACTGACTCTGA 1412  
  
QY 5766 TTATTTCTGACCTGTTTAAAGCTTTAGGGAAGAGCTTTAGAGAAAGAACCGTAGAGATCAT 5825  
DB 1413 TTATTTCTGACCTGTTTAAAGCTTTAGGGAAGAGCTTTAGAGAAAGAACCGTAGAGATCAT 1472  
  
QY 5826 TCTTTTCGTCACCTTTGTGTATCTTTCAATAATTTCTTCTTGGAAACAGCTATTTTTCGATT 5885  
DB 1473 TCTTTTCGTCACCTTTGTGTATCTTTCAATAATTTCTTCTTGGAAACAGCTATTTTTCGATT 1532  
  
QY 5886 TGCAAGAAATTCGACAGAGCAAGAGAGGTGGCACAATCTCTCTCAGGATGCAAGAAAT 5945  
DB 1533 TGCAAGAAATTCGACAGAGCAAGAGAGGTGGCACAATCTCTCTCAGGATGCAAGAAAT 1592  
  
QY 5946 TATTGACACTAAATTTGAAGCTGATCAGCTCATCAGTGGGAGAGCTGGAAGCTGAAGGA 6005  
DB 1593 TATTGACACTAAATTTGAAGCTGATCAGCTCATCAGTGGGAGAGCTGGAAGCTGAAGGA 1652  
  
QY 6006 CGAGAGATTAACGCG----- 6020  
DB 1653 CGAGAGATTAACGCGGATCTCTTCTTTTCTTGTAGCGCAAGAGATTTTGTGATCCA 1712  
  
QY 6021 -----GCCAATAAGCAAGAAATCTCTCTGCAACATGT 6053  
DB 1713 GTTACTTACTTATAGAAATCCATCAAGCCCAATAGCAAGAAATCTCTCTGCAACATGT 1772  
  
QY 6054 TGAAGAGCTTTGCAACAAACAAACCTAAAGTTTCAAGAAATTTTCGGAATTCACAAA 6113  
DB 1773 TGAAGAGCTTTGCAACAAACAAACCTAAAGTTTCAAGAAATTTTCGGAATTCACAAA 1832  
  
QY 6114 ATTTCTTCAGGATCTTTCTTCACTGATGCTGATCTGCTTGGAAATAGAGCAAAACCG 6173  
DB 1833 ATTTCTTCAGGATCTTTCTTCACTGATGCTGATCTGCTTGGAAATAGAGCAAAACCG 1892  
  
QY 6174 TTTTCCCAACATAAAACCATATAATAATAACAGGTAAAGCTGATAGTGCAGCTAG 6233  
DB 1893 CTTTCCCAACATAAAACCATATAATAATAACAGGTAAAGCTGATAGTGCAGCTAG 1949  
  
QY 6234 TGTTCAGGTTCCGATTAATTAATGCGCAGCTATATTTCTGCTTATTTATGTCNAATGA 6293



Db 1950 TGTTCAGGTTGGATTATATTAATGCCAGCTATATTTCTGGTATTTATGTCCTCAATGA 2009  
Qy 6294 ATTATTTGCTACTCAAGTCCACTACACAGAACAGTTGGAGATTTTGGAGAAATGGTGTG 6353  
Db 2010 ATTATTTGCTACTCAAGTCCACTACACAGAACAGTTGGAGATTTTGGAGAAATGGTGTG 2069  
Qy 6354 GGAACACAGGGCAAAACATTAGTAATGCTAACACAGTGTTTTGAAGAGGACGGATCAG 6413  
Db 2070 GGAACACAGGCAAAACATTAGTAATGCTAACACAGTGTTTTGAAGAGGACGGATCAG 2129  
Qy 6414 ATGCCATCAGTATTTGCCACAGGACCAACAGGACGTTACTGTCTTTTGGAGATATAGTGAT 6473  
Db 2130 ATGCCATCAGTATTTGCCACAGGACCAACAGGACGTTACTGTCTTTTGGAGATATAGTGAT 2189  
Qy 6474 TACAAAGCTAATGGAGGATTTCAAAATAGATTGGACTATCAGGATCTGAAAATTTGAAG 6533  
Db 2190 TACAAAGCTAATGGAGGATTTCAAAATAGATTGGACTATCAGGATCTGAAAATTTGAAG 2249  
Qy 6534 GCATGGGATTCATGACTGTTTCGACAGTGAATCTTACTGCTGGCCAGACATGGGCT 6593  
Db 2250 GCATGGGATTCATGACTGTTTCGACAGTGAATCTTACTGCTGGCCAGACATGGGCT 2309  
Qy 6594 TCTGAGAACAGCGCCCTCTAAATTCATTTGTGAAGTTGGTTTCGAGCAAGCAGGCGACA 6653  
Db 2310 TCTGAGAACAGCGCCCTCTAAATTCATTTGTGAAGTTGGTTTCGAGCAAGCAGGCGACA 2369  
Qy 6654 TGACACCAACCTATGATTTGTCACCTGAGTGTGGAGTTGGAAGAACTGGAGTTTAT 6713  
Db 2370 TGACACCAACCTATGATTTGTCACCTGAGTGTGGAGTTGGAAGAACTGGAGTTTAT 2429  
Qy 6714 TGTCTGGACCAATTAACACACATATAATGACCATGATTTTGTGGATATATATGACT 6773  
Db 2430 TGTCTGGACCAATTAACACACATATAATGACCATGATTTTGTGGATATATATGACT 2489  
Qy 6774 AGTAGCTGAACCTGAGAAAGTGAAGATGTGCATGTGTGCAGAAATCTGGCAACATATCTT 6833  
Db 2490 AGTAGCTGAACCTGAGAAAGTGAAGATGTGCATGTGTGCAGAAATCTGGCAACATATCTT 2549  
Qy 6834 TTTACACCAAGTGTCTGATCTTATCAATAAGGAAAGTAATCAGCCCATCTGTTT 6893  
Db 2550 TTTACACCAAGTGTCTGATCTTATCAATAAGGAAAGTAATCAGCCCATCTGTTT 2609  
Qy 6894 TGTAACTATTCAGCACTTCAGAGATGACTCTTTGGACGCCATGGAGGATGATGTGA 6953  
Db 2610 TGTAACTATTCAGCACTTCAGAGATGACTCTTTGGACGCCATGGAGGATGATGTGA 2669  
Qy 6954 GCTTGAATGGGAAGAAACCACTATGTAAATATTCAGACCAAGGATACAAATTTGGAAGAGA 7013  
Db 2670 GCTTGAATGGGAAGAAACCACTATGTAAATATTCAGACCAAGGATACAAATTTGGAAGAGA 2729  
Qy 7014 TTTTAAATCCAGGGGCAAGTATACCCCTCATCTTCCGAATTTGGAATGTGAACCT 7073  
Db 2730 TTTTAAATCCAGGGGCAAGTATACCCCTCATCTTCCGAATTTGGAATGTGAACCT 2789  
Qy 7074 TAAAGAAATCTATGCTTCTCTCACTGTCCTTT 7108  
Db 2790 TAAAGAAATCTATGCTTCTCTCACTGTGCTTT 2824

RESULT 4  
US-09-361-096A-21  
; Sequence 21, Application US/09361096A  
; Patent No. 6492495  
; GENERAL INFORMATION:  
; APPLICANT: MOLLER, NIELS P.H.  
; APPLICANT: MOLLER, KARIN B.  
; APPLICANT: ULLRICH, AXEL  
; TITLE OF INVENTION: PTP-831: A NOVEL PROTEIN TYROSINE PHOSPHATASE  
; FILE REFERENCE: 038602/0686  
; CURRENT APPLICATION NUMBER: US/09/361,096A  
; CURRENT FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: 08/449,609

; PRIOR FILING DATE: 1995-05-24  
; PRIOR APPLICATION NUMBER: 08/036,210  
; PRIOR FILING DATE: 1995-03-23  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 21  
; LENGTH: 3973  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PTP831-RD#2  
; NAME/KEY: modified base  
; LOCATION: (1)..(3973)  
; OTHER INFORMATION: n = unknown nucleotide  
; NAME/KEY: CDS  
; LOCATION: (1)..(2694)  
; US-09-361-096A-21

Query Match 36.8%; Score 2618.6; DB 4; Length 3973;  
Best Local Similarity 97.2%; Pred. No. 0;  
Matches 2716; Conservative 0; Mismatches 4; Indels 75; Gaps 2;  
Qy 4386 AACAGTCCCAAGTGTCCCAAAATATTGCTTTTCTGATGTCAGTCAACTAGTGCAC 4445  
Db 33 AAAAGTCCCAAGTGTCCCAAAATATTGCTTTTCTGATGTCAGTCAACTAGTGCAC 92  
Qy 4446 ATTGATGGATTAAGACCTGACATCTCTGGCTACTTTCAAAATTACAAATTTACCC 4505  
Db 93 ATTGATGGATTAAGACCTGACATCTCTGGCTACTTTCAAAATTACAAATTTACCC 152  
Qy 4506 TCAACTCTGCTGCTCAAAATGCGAAATGCGAAATGCGAAATGCGAAATGCGAAAT 4565  
Db 153 TCAACTCTGCTGCTCAAAATGCGAAATGCGAAATGCGAAATGCGAAATGCGAAAT 212  
Qy 4566 AATTCAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGAATTAAGAA 4625  
Db 213 AATTCAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGAATTAAGAA 272  
Qy 4626 TAGATGGTATAGATTTCCAGTGGCTGCCAGACCAATGCTGGCTATGCGAATGCTTCAA 4685  
Db 273 TAGATGGTATAGATTTCCAGTGGCTGCCAGACCAATGCTGGCTATGCGAATGCTTCAA 332  
Qy 4686 CTGGATTTCTACAAAACTCTGCTGGCTGCCAGTGGCTGCCAGTGGCTGCCAGTGGCT 4745  
Db 333 CTGGATTTCTACAAAACTCTGCTGGCTGCCAGTGGCTGCCAGTGGCTGCCAGTGGCT 392  
Qy 4746 AGTAGCAACATCACTTTTAGCATCAGCATAGCTGGAGTGAACCTGCTGTCTATCTTGG 4805  
Db 393 AGTAGCAACATCACTTTTAGCATCAGCATAGCTGGAGTGAACCTGCTGTCTATCTTGG 452  
Qy 4806 ACCAATGTTTATCTGATGATGTCAAATCGTAGATTAATGATGAATTTATATATCTT 4865  
Db 453 ACCAATGTTTATCTGATGATGTCAAATCGTAGATTAATGATGAATTTATATATCTT 512  
Qy 4866 CATCAAGTCAATGAAGAAAAATAAACAATGAAGATTTAGAAATTTAGAAATTTACAAG 4925  
Db 513 CATCAAGTCAATGAAGAAAAATAAACAATGAAGATTTAGAAATTTAGAAATTTACAAG 572  
Qy 4926 GTATTTCTGTAGTATCACTGCAATTTACTGGGAACATTAAGTGTGTCATATGTAGAGG 4985  
Db 573 GTATTTCTGTAGTATCACTGCAATTTACTGGGAACATTAAGTGTGTCATATGTAGAGG 632  
Qy 4986 GTCAAGTGTGAATGATTTGTTACTTACTTTAGAAATCAGCCCCCAAGGCCACCTAACAA 5045  
Db 633 GTCAAGTGTGAATGATTTGTTACTTACTTTAGAAATCAGCCCCCAAGGCCACCTAACAA 692  
Qy 5046 CATGACATTTTCAAGAGATACAGATGAAGTTTACAAATTTCAATTAACCTTCTCTCTCC 5105  
Db 693 CATGACATTTTCAAGAGATACAGATGAAGTTTACAAATTTCAATTAACCTTCTCTCTCC 752  
Qy 5106 TTCTCAACCTAATGGAATATCCAGTATATCAAGCTCTGGTTTACCGAAGAGATGATCC 5165  
Db 753 TTCTCAACCTAATGGAATATCCAGTATATCAAGCTCTGGTTTACCGAAGAGATGATCC 812





STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449,609  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/036,210  
FILING DATE: 23-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7683-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2692 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2149  
OTHER INFORMATION: /note= "N=x-unknown nucleotide"  
US-08-449-609-14

Query Match 15.7%; Score 1118.8; DB 2; Length 2692;  
Best Local Similarity 93.7%; Pred. No. 6.6e-289;  
Matches 1218; Conservative 0; Mismatches 7; Indels 75; Gaps 2;  
QY 5881 GCATTTGCAAGATTTCGACAGAGCAGAGAGAGGTTGGCACAATCTCTCTCAGGATGCA 5940  
DB 247 GCTTTGTTTGAATTCGACAGAGCAGAGAGAGGTTGGCACAATCTCTCTCAGGATGCA 306  
QY 5941 GAAATTAATGACACTAAATTTGAAGCTGGATCAGCTCATCAGTGGCAGACCTGGAACTG 6000  
DB 307 GAAATTAATGACACTAAATTTGAAGCTGGATCAGCTCATCAGTGGCAGACCTGGAACTG 366  
QY 6001 AAGGACGAGATTACGCG----- 6020  
DB 367 AAGGACGAGATTACGCGGATCTCTCATTTTCTTTAGACGCAAGGAGATTTTGTGTC 426  
QY 6021 -----GCCAATAAGCAAGAAATCTCTCTGCAA 6048  
DB 427 ATCCAGTTACTTATAGAAATCCATCAAGCCATTAAGCAAGAAATCTCTCTGCAA 486  
QY 6049 CATGTTGAAGAGCTTTGCAACAAACAACTTAAGTTTCAAGAGAAATTTTCGGAATTA 6108  
DB 487 CATGTTGAAGAGCTTTGCAACAAACAACTTAAGTTTCAAGAGAAATTTTCGGAATTA 546  
QY 6109 CCAAAATTTCTCAGATCTTTCTTCACTGATGCTGATCTGCTTTGGAATAGACAAA 6168  
DB 547 CCAAAATTTCTCAGATCTTTCTTCACTGATGCTGATCTGCTTTGGAATAGACAAA 606  
QY 6169 AACCGTTTCCAAACATAAACCATATAATAATAACAGAGTAAAGCTGATAGCTGAC 6228  
DB 607 AACCGTTTCCAAACATAAACCATATAATAATAACAGAGTAAAGCTGATAGCTGAC 663  
QY 6229 GCTAGTGTTCAGGTTCCGATATATTAATGCGAGCTATATTTCTGGTATTATGTCAC 6288  
DB 664 GCTAGTGTTCAGGTTCCGATATATTAATGCGAGCTATATTTCTGGTATTATGTCAC 723

QY 6289 AATGAATTTATTGCTACTCAAGGTCCACTACAGGAAACAGTTGGAGATTTTGGAGAAATG 6348  
DB 724 AATGAATTTATTGCTACTCAAGGTCCACTACAGGAAACAGTTGGAGATTTTGGAGAAATG 783  
QY 6349 GTGTGGAAACAGGCAAAACATTTAGTAACTTAACACAGTGTGTTTGAAGAGCAGCG 6408  
DB 784 GTGTGGAAACAGGCAAAACATTTAGTAACTTAACACAGTGTGTTTGAAGAGCAGCG 843  
QY 6409 ATCAGATGTCATCAGTATTGGCCAGAGGACAAACAGCCAGTTACTCTCTTTGGAGATATA 6468  
DB 844 ATCAGATGTCATCAGTATTGGCCAGAGGACAAACAGCCAGTTACTCTCTTTGGAGATATA 903  
QY 6469 GTGATTAACAAGCTAATGAGGATGTTCAAAATAGATTTGACATATCAGGATCTGAAAAAT 6528  
DB 904 GTGATTAACAAGCTAATGAGGATGTTCAAAATAGATTTGACATATCAGGATCTGAAAAAT 963  
QY 6529 GAAAGGCATGGGATTCATGACTGTTCCACAGTGTAACTTTACTGCTGCGCCAGAGCAT 6588  
DB 964 GAAAGGCATGGGATTCATGACTGTTCCACAGTGTAACTTTACTGCTGCGCCAGAGCAT 1023  
QY 6589 GGGGTTCTTGAGAACAGCGCCCTCTAAATTTCACTTTGCAAGTTGTTTGCAGCAAGCAGG 6648  
DB 1024 GGGGTTCTTGAGAACAGCGCCCTCTAAATTTCACTTTGCAAGTTGTTTGCAGCAAGCAGG 1083  
QY 6649 GCACATGACACCACTATGATTTGTTCACTGAGTGTCTGAGTTGGAGAACTGGAGTT 6708  
DB 1084 GCACATGACACCACTATGATTTGTTCACTGAGTGTCTGAGTTGGAGAACTGGAGTT 1143  
QY 6709 TTTATTTGCTCTGAGCCATTTAAACAAATATAAATGACATGATTTTGTGGATATATAT 6768  
DB 1144 TTTATTTGCTCTGAGCCATTTAAACAAATATAAATGACATGATTTTGTGGATATATAT 1203  
QY 6769 GGACTAGTAGCTGAACTGAGAGTGAAGAAATGTCATGTCAGATCTGGCAGAACTGGCAGCAT 6828  
DB 1204 GGACTAGTAGCTGAACTGAGAGTGAAGAAATGTCATGTCAGATCTGGCAGAACTGGCAGCAT 1263  
QY 6829 ATCTTTTACACAGTGCATTTCTGGATCTCTTATCAAAATAGGAAAGTAAATCAGCCCATC 6888  
DB 1264 ATCTTTTACACAGTGCATTTCTGGATCTCTTATCAAAATAGGAAAGTAAATCAGCCCATC 1323  
QY 6889 TGTTTTGTTAACTATTTCAGCATTTCAGAAAGATGACTCTTTGGACGCCATGGAGGTGAT 6948  
DB 1324 TGTTTTGTTAACTATTTCAGCATTTCAGAAAGATGACTCTTTGGACGCCATGGAGGTGAT 1383  
QY 6949 GTTGAGCTTGAATGGGAGAAACCACTATGTAATATTTCAGACCAAGGATCAATTTGGA 7008  
DB 1384 GTTGAGCTTGAATGGGAGAAACCACTATGTAATATTTCAGACCAAGGATCAATTTGGA 1443  
QY 7009 AGAGATTTTAAATCCAGGGGCCAAAGTTTACCCCTCTATTCTTCGAAATGAAATGTGC 7068  
DB 1444 AGAGATTTTAAATCCAGGGGCCAAAGTTTACCCCTCTATTCTTCGAAATGAAATGTGC 1503  
QY 7069 AACCTTAAAGAAATATCTATGCTTTCTCTCAGTGTGCTTT 7108  
DB 1504 AACCTTAAAGAAATATCTATGCTTTCTCTCAGTGTGCTTT 1543

RESULT 7  
US-09-361-096A-14  
; Sequence 14, Application US/09361096A  
; Patent No. 6492495  
; GENERAL INFORMATION:  
; APPLICANT: MOLLER, NIELS P.H.  
; APPLICANT: MOLLER, KARIN B.  
; APPLICANT: ULLRICH, AXEL  
; TITLE OF INVENTION: PTP-531: A NOVEL PROTEIN TYROSINE PHOSPHATASE  
; FILE REFERENCE: 038602/0686  
; CURRENT APPLICATION NUMBER: US/09/361,096A  
; CURRENT FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: 08/449,609  
; PRIOR FILING DATE: 1995-05-24  
; PRIOR APPLICATION NUMBER: 08/036,210

;; PRIOR FILING DATE: 1995-03-23  
;; NUMBER OF SEQ ID NOS: 53  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 14  
;; LENGTH: 2692

;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: modified base  
;; LOCATION: (1)..(2692)  
;; OTHER INFORMATION: n = unknown nucleotide  
;; NAME/KEY: CDS  
;; LOCATION: (92..139, 259..1414)  
US-09-361-096A-14

Query Match 15.7%; Score 1118.8; DB 4; Length 2692;

Best Local Similarity 93.7%; Pred. No. 6.6e-289;  
Matches 1218; Conservative 0; Mismatches 7; Indels 75; Gaps 2;

|    |      |   |      |
|----|------|---|------|
| QY | 5881 | GCATTTGCAAGAAATTCGACAGAGCAGAAAGAGTGGGCACATACCTCTCCTCAGGATGCA    | 5940 |
| Db | 247  | GCTTGTTTGAATTCGACAGAGCAGAAAGAGTGGGCACATACCTCTCCTCAGGATGCA       | 306  |
| QY | 5941 | GAATTAATTGACACTAAATTTGAAGCTGGATCAGCTCATCACTAGTGGCAGACCTGGAAC    | 6000 |
| Db | 307  | GAATTAATTGACACTAAATTTGAAGCTGGATCAGCTCATCACTAGTGGCAGACCTGGAAC    | 366  |
| QY | 6001 | AAGGACGAGAGATTAACGGC-----                                       | 6020 |
| Db | 367  | AAGGACGAGAGATTAACGGCATACTCTCATTTTCTTTAGACGCAAGGAGATTTTGTG       | 426  |
| QY | 6021 | -----GCCAATAAGCAAGAAATCCCTCTCTCTGCAA                            | 6048 |
| Db | 427  | ATCCAGTTACTAGTTATAGAAATCCATCAAGCCCAATAGCAAGAAATCCCTCTCTGCAA     | 486  |
| QY | 6049 | CATGTTGAAGAGCTTTGCAACAAACCACTAAAGTTTCAAGAGAAATTTTCGGAATTA       | 6108 |
| Db | 487  | CATGTTGAAGAGCTTTGCAACAAACCACTAAAGTTTCAAGAGAAATTTTCGGAATTA       | 546  |
| QY | 6109 | CCAAATTTCTCAGATCTTTCTCAACTGATGCTGATCTGCTTGGATAGACAAA            | 6168 |
| Db | 547  | CCAAATTTCTCAGATCTTTCTCAACTGATGCTGATCTGCTTGGATAGACAAA            | 606  |
| QY | 6169 | AACGGTTTCCCAACATAAAACCATATAATAATAACAGAGTAAAGCTAGTAGTAC          | 6228 |
| Db | 607  | AACGGTTTCCCAACATAAAACCAT---ATAATAATAACAGAGTAAAGCTAGTAGTAC       | 663  |
| QY | 6229 | GCTAGTGTTCAGGTTCCGATTATATAATGCGAGCTATATTTCTGTTTATTTATGTCCA      | 6288 |
| Db | 664  | GCTAGTGTTCAGGTTCCGATTATATAATGCGAGCTATATTTCTGTTTATTTATGTCCA      | 723  |
| QY | 6289 | AATGAATTTATGCTACTCAAGTCCACTACAGGAAACAGTTGGAGATTTTGGAGATG        | 6348 |
| Db | 724  | AATGAATTTATGCTACTCAAGTCCACTACAGGAAACAGTTGGAGATTTTGGAGATG        | 783  |
| QY | 6349 | GTGTGGAAACAGGCAAAACATTAGTAACTGCTTAACACAGCTTTTGAAGAGACGG         | 6408 |
| Db | 784  | GTGTGGAAACAGGCAAAACATTAGTAACTGCTTAACACAGCTTTTGAAGAGACGG         | 843  |
| QY | 6409 | ATCAGATGCCATCAGTATTGGCCAGAGGCAACACAGCCAGTTACTGTCTTTGGAGATATA    | 6468 |
| Db | 844  | ATCAGATGCCATCAGTATTGGCCAGAGGCAACACAGCCAGTTACTGTCTTTGGAGATATA    | 903  |
| QY | 6469 | GTGATTAACAAGCTAATGAGGATGTTCAAAATAGATGGAATTCAGGGATCTGAAATTT      | 6528 |
| Db | 904  | GTGATTAACAAGCTAATGAGGATGTTCAAAATAGATGGAATTCAGGGATCTGAAATTT      | 963  |
| QY | 6529 | GAAAGCATGGGATTCAGTACTGTTCCACAGTGTAACTTTACTGCTTGGCCAGAGCAT       | 6588 |
| Db | 964  | GAAAGCATGGGATTCAGTACTGTTCCACAGTGTAACTTTACTGCTTGGCCAGAGCAT       | 1023 |
| QY | 6589 | GGGGTTCTCTGAGAACAGCGCCCTCTAAATTTCACTTTGTGAAGTGTGGTTCGAGCAAGCAGG | 6648 |

|    |      |   |      |
|----|------|---|------|
| Db | 1024 | GGGGTTCTCTGAGAACAGCGCCCTCTAAATTCATTTGTGAAGTGTGGTTTCGAGCAAGCAGG  | 1083 |
| QY | 6649 | GCACATGACACACACCTATGATTTCTCACTGCACTGCTGAGTTGGAGAACTCGAGTT       | 6708 |
| Db | 1084 | GCACATGACACACACCTATGATTTCTCACTGCACTGCTGAGTTGGAGAACTCGAGTT       | 1143 |
| QY | 6709 | TTTATTGCTCTGAGACCATTTAAACACACATATAAATGACCATGATTTTGTGGATATAT     | 6768 |
| Db | 1144 | TTTATTGCTCTGAGACCATTTAAACACACATATAAATGACCATGATTTTGTGGATATAT     | 1203 |
| QY | 6769 | GGACTAGTAGCTGAACCTGAGAAAGTGAAGATGTCATGTCGACGAAATCTGCGACAGTAT    | 6828 |
| Db | 1204 | GGACTAGTAGCTGAACCTGAGAAAGTGAAGATGTCATGTCGACGAAATCTGCGACAGTAT    | 1263 |
| QY | 6829 | ATCTTTTACACCACTGATCTCTTATCAATTAAGGGAAGTAATCAGCCCATC             | 6888 |
| Db | 1264 | ATCTTTTACACCACTGATCTCTTATCAATTAAGGGAAGTAATCAGCCCATC             | 1323 |
| QY | 6889 | TGTTTGTGTTAACTATTCAGCACTTCAGAAAGATGGAATCTTTTGGACGCCATGGAAGGTGAT | 6948 |
| Db | 1324 | TGTTTGTGTTAACTATTCAGCACTTCAGAAAGATGGAATCTTTTGGACGCCATGGAAGGTGAT | 1383 |
| QY | 6949 | GTTGAGCTTGAATGGGAAAGAACCACTATGTAAATATTTCAGACCAAGAGATCAATTTGA    | 7008 |
| Db | 1384 | GTTGAGCTTGAATGGGAAAGAACCACTATGTAAATATTTCAGACCAAGAGATCAATTTGA    | 1443 |
| QY | 7009 | AGAGATTTTAAATCCAGGGCCAAAGTTACCCCTCATCTTCCGAAATGAAATGTGC         | 7068 |
| Db | 1444 | AGAGATTTTAAATCCAGGGCCAAAGTTACCCCTCATCTTCCGAAATGAAATGTGC         | 1503 |
| QY | 7069 | AACCTTAAAGAAATATCTATGCTTCTCTCAGTGTGCTTT                         | 7108 |
| Db | 1504 | AACCTTAAAGAAATATCTATGCTTCTCTCAGTGTGCTTT                         | 1543 |

## RESULT 8

US-08-036-210-10  
; Sequence 10, Application US/08036210  
; Patent No. 5585233  
; GENERAL INFORMATION:  
; APPLICANT: Moller, Niels P.H.  
; APPLICANT: Moller, Karin B.  
; APPLICANT: Ullrich, Axel  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASE  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: PENNIE & EDMONDS  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/036,210  
; FILING DATE: 23-MAR-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistrick, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7683-025  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2309 base pairs

```
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1766
; OTHER INFORMATION: /note= "N=x-unknown nucleotide"
US-08-036-210-10

Query Match      15.1%; Score 1071.6; DB 1; Length 2309;
Best Local Similarity 99.4%; Pred. No. 2.7e-276;
Matches 1087; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 6015 AACGGGCCAATAAGCAAGAAATCTTCTCGCAACATGTTGGAAGAGCTTTGCAACAACAA 6074
DB 70 AATGAGGCCAATAAGCAAGAAATCTTCTCGCAACATGTTGGAAGAGCTTTGCAACAACAA 129
QY 6075 CAACCTAAAGTTTCAAGAGAAATTTTCGGAAATACCAAAATTTCTTCAGGATCTTTCTTC 6134
DB 130 CAACCTAAAGTTTCAAGAGAAATTTTCGGAAATACCAAAATTTCTTCAGGATCTTTCTTC 189
QY 6135 AACTGATGCTGATCTGCTTGGAAATAGACCAAAACCGTTTCCCAAAATATAAACCATA 6194
DB 190 AACTGATGCTGATCTGCTTGGAAATAGACCAAAACCGTTTCCCAAAATATAAACCATA 248
QY 6195 TAATAATAATAACAGAGTAAGCTGATGCTGACGCTAGTGTTCAGGTTTCGGAATATAT 6254
DB 249 --ATAATAATAACAGAGTAAGCTGATGCTGACGCTAGTGTTCAGGTTTCGGAATATAT 306
QY 6255 TAATGCCAGCTATATTTCTGTTATTTATGTCGAAATGAATTTATGCTACTCAAGGTC 6314
DB 307 TAATGCCAGCTATATTTCTGTTATTTATGTCGAAATGAATTTATGCTACTCAAGGTC 366
QY 6315 ACTACCAAGCAAGTGTGGAGATTTTGGAGATGTTGGAGATGTTGGAGATGTTGGAGAT 6374
DB 367 ACTACCAAGCAAGTGTGGAGATTTTGGAGATGTTGGAGATGTTGGAGATGTTGGAGAT 426
QY 6375 AGTAATGCTAAACACAGTGTGTTGAAAGAGCGATGATGCTAGTGTGCTGATTTGGCCAG 6434
DB 427 AGTAATGCTAAACACAGTGTGTTGAAAGAGCGATGATGCTAGTGTGCTGATTTGGCCAG 486
QY 6435 GGACAAACAGCCAGTGTGCTTTGGAGATATAGTATTAACAACTAATGAGAGATGT 6494
DB 487 GGACAAACAGCCAGTGTGCTTTGGAGATATAGTATTAACAACTAATGAGAGATGT 546
QY 6495 TCAATAGATGACATCTAGGATCTGAAATTTGAAAGCGATGCGGATTTGCTGATGCTGT 6554
DB 547 TCAATAGATGACATCTAGGATCTGAAATTTGAAAGCGATGCGGATTTGCTGATGCTGT 606
QY 6555 TCGACAGTGTAACTTTTACTGCTGCGCAGAGCATGCGGTTCTCTGAGAACAGCGCCCTCT 6614
DB 607 TCGACAGTGTAACTTTTACTGCTGCGCAGAGCATGCGGTTCTCTGAGAACAGCGCCCTCT 666
QY 6615 AATTCACTTTGGAAGTTGTTGAGCAAGAGCGGCAATGACCAACCACTATGATTTGT 6674
DB 667 AATTCACTTTGGAAGTTGTTGAGCAAGAGCGGCAATGACCAACCACTATGATTTGT 726
QY 6675 TCACTGCTGCTGGAAGTTGGAAGACTGGAGTTTATTTGCTCTGACCAATTAACACA 6734
DB 727 TCACTGCTGCTGGAAGTTGGAAGACTGGAGTTTATTTGCTCTGACCAATTAACACA 786
QY 6735 ACATATAAATGACCATGATTTTGTGGATATATATGACTAGTAGCTGAAGTGAAGTGA 6794
DB 787 ACATATAAATGACCATGATTTTGTGGATATATATGACTAGTAGCTGAAGTGAAGTGA 846
QY 6795 AAGAATGTGCATGGTCCAGAAATCTGGCACAGTATATCTTTTACACAGTGAATTTGGA 6854
DB 847 AAGAATGTGCATGGTCCAGAAATCTGGCACAGTATATCTTTTACACAGTGAATTTGGA 906
QY 6855 TCTCTTATCAATAAGGGAAGTAAATCAGGCCATCTGTTTGTAACTATTCAGACTTCA 6914
DB 907 TCTCTTATCAATAAGGGAAGTAAATCAGGCCATCTGTTTGTAACTATTCAGACTTCA 966
```

```
QY 6915 GAAGATGGAGTCTTTGGACGCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGAAACCAC 6974
DB 967 GAAGATGGAGTCTTTGGACGCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGAAACCAC 1026
QY 6975 TATGTAATAATTTTCAGACCAAGGATACATTCGGAAGATTTTAAATCCAGGGGCCAA 7034
DB 1027 TATGTAATAATTTTCAGACCAAGGATACATTCGGAAGATTTTAAATCCAGGGGCCAA 1086
QY 7035 AGTTACCCCTCATTTCTCCGAATTTGAAATGTGCAACCTTTAAAGAAATATCTATGCTTCT 7094
DB 1087 AGTTACCCCTCATTTCTCCGAATTTGAAATGTGCAACCTTTAAAGAAATATCTATGCTTCT 1146
QY 7095 CTCACGTGCTCTTT 7108
DB 1147 CTCACGTGCTCTTT 1160
```

```
RESULT 9
US-08-449-609-10
; Sequence 10, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSES: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,609
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/036,210
; FILING DATE: 23-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; NAME: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2309 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1766
; OTHER INFORMATION: /note= "N=x-unknown nucleotide"
US-08-449-609-10
```

```
Query Match      15.1%; Score 1071.6; DB 2; Length 2309;
Best Local Similarity 99.4%; Pred. No. 2.7e-276;
Matches 1087; Conservative 0; Mismatches 4; Indels 3; Gaps 1;
```





Db 547 TCAATAGATTGGACTATCAGGATCTGAAATTTGAAAGCATGGGATTGCATGACTGT 606  
Qy 6555 TCGACAGTGAATCTTACTCCCTGGCCAGAGCATGGGGTTCTTGAGAACAGCCCTCT 6614  
Db 607 TCGACAGTGAATCTTACTCCCTGGCCAGAGCATGGGGTTCTTGAGAACAGCCCTCT 666  
Qy 6615 AATTCATTTGTGAAGTTGGTTCGAGCAAGCAGGCGACATGACACACCACTATGTTGT 6674  
Db 667 AATTCATTTGTGAAGTTGGTTCGAGCAAGCAGGCGACATGACACACCACTATGTTGT 726  
Qy 6675 TCACTGAGTGGTGGAGTTCGAGAACTGGAGTATTTTATGCTCTGGACCATTTAACACA 6734  
Db 727 TCACTGAGTGGTGGAGTTCGAGAACTGGAGTATTTTATGCTCTGGACCATTTAACACA 786  
Qy 6735 ACATATAAATGACCATGATTTTGTGGATATATATGAGCTAGTAGTGTGAACTGAGAACTGA 6794  
Db 787 ACATATAAATGACCATGATTTTGTGGATATATATGAGCTAGTAGTGTGAACTGAGAACTGA 846  
Qy 6795 AGAATGTGATGGTGCAGATCTGGCAGATATATCTTTTACACCATGCTATCTGGA 6854  
Db 847 AGAATGTGATGGTGCAGATCTGGCAGATATATCTTTTACACCATGCTATCTGGA 906  
Qy 6855 TCTCTTATCAATAAGGAGTAAATCAGCCCATCTGTTTGTGTTAACTATTTCAGCACTTCA 6914  
Db 907 TCTCTTATCAATAAGGAGTAAATCAGCCCATCTGTTTGTGTTAACTATTTCAGCACTTCA 966  
Qy 6915 GAAGATGGACTCTTTTGGACCCCATGGAAGTGTGTTGAGCTTTGAATGGGAAGAACCCAC 6974  
Db 967 GAAGATGGACTCTTTTGGACCCCATGGAAGTGTGTTGAGCTTTGAATGGGAAGAACCCAC 1026  
Qy 6975 TATGTAATAATTCAGACCAAGATACAAATGGAAGATTTTAAATCCAGGGGCCAA 7034  
Db 1027 TATGTAATAATTCAGACCAAGATACAAATGGAAGATTTTAAATCCAGGGGCCAA 1086  
Qy 7035 AGTTACCCCTCATCTTCCGAATTTGAATGTGCAACCTTAAAGAAATATCTATGCTTCT 7094  
Db 1087 AGTTACCCCTCATCTTCCGAATTTGAATGTGCAACCTTAAAGAAATATCTATGCTTCT 1146  
Qy 7095 CTCACGTGCTTT 7108  
Db 1147 CTCACGTGCTTT 1160

## RESULT 11

US-08-036-210-6  
; Sequence 6, Application US/08036210  
; Patent No. 5585233  
; GENERAL INFORMATION:  
; APPLICANT: Moller, Niels P.H.  
; APPLICANT: Moller, Karin B.  
; APPLICANT: Ullrich, Axel  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASE  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/036,210  
; FILING DATE: 23-MAR-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misbrock, S. Leslie  
; REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7683-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864/9741  
TELEX: 66141 PENNIS  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2173 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1630  
OTHER INFORMATION: /note= "N=x=unknown nucleotide"  
US-08-036-210-6

Query Match 11.1%; Score 789.6; DB 1; Length 2173;  
Best Local Similarity 86.9%; Pred. No. 9.8e-201;  
Matches 951; Conservative 0; Mismatches 4; Indels 139; Gaps 2;  
Qy 6015 AAGCGGCGCAATAAGCAAGAAATCTTCTGCAACATGTTGAGAGCTTTGCAACAACAA 6074  
Db 70 AATGAGGCCAATAAGCAAGAAATCTTCTGCAACATGTTGAGAGCTTTGCAACAACAA 129  
Qy 6075 CAACCTAAAGTTTCAAGAAAGAAATTTTCGGAATTTACCAAAATTTCTTCAGGATCTTTCTTC 6134  
Db 130 CAACCTAAAGTTTCAAGAAAGAAATTTTCGGAATTTACCAAAATTTCTTCAGGATCTTTCTTC 189  
Qy 6135 AACTGATGCTGATCTGCTTGGAAATAGAGCAAAAAACCGTTTCCCAAAACATAAAACCATTA 6194  
Db 190 AACTGATGCTGATCTGCTTGGAAATAGAGCAAAAAACCGTTTCCCAAAACATAAAACCAT - 248  
Qy 6195 TAATAATAATACAGAGTAAAGCTGATAGCTGACGCTAGTGTTCAGAGTTCCGATTTATAT 6254  
Db 249 --ATAATAATACAGAGTAAAGCTGATAGCTGACGCTAGTGTTCAGAGTTCCGATTTATAT 306  
Qy 6255 TAATGCCAGCTATATTTCTGGTTATTTATGTCCAAATGAATTTATTTGCTACTCAAGGTCC 6314  
Db 307 TAATGCCAGCTATATTTCTGGTTATTTATGTCCAAATGAATTTATTTGCTACTCAAGGTCC 366  
Qy 6315 ACTACCAAGAAACAGTTGGAGATTTTGGAGATGTTGTGGGAAACAGGCGCAAAAAACATT 6374  
Db 367 ACTACCAAGAAACAGTTGGAGATTTTGGAGATGTTGTGGGAAACAGGCGCAAAAAACATT 426  
Qy 6375 AGTAATGCTAACACAGTGTTTTGAAGAAAGGAGCGATCAGATGCCATCAGTATTGGCCAGA 6434  
Db 427 AGTAATGCTAACACAGTGTTTTGAAGAAAGGAGCGATCAGATGCCATCAGTATTGGCCAGA 486  
Qy 6435 GGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGAATTACAAAGCTAATGGAGGATGT 6494  
Db 487 GGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGAATTACAAAGCTAATGGAGGATGT 546  
Qy 6495 TCAATAGATTGGACTATCAGGGATCTGAAATTTGAAAGGCATGGGATTTGCATGCTGT 6554  
Db 547 TCAATAGATTGGACTATCAGGGATCTGAAATTTGAAAGGCATGGGATTTGCATGCTGT 606  
Qy 6555 TCGACAGTGTAACTTTTACTGCTGCGCCAGAGCATGGGTTTCTTGAGAAACAGCGCCCTCT 6614  
Db 607 TCGACAGTGTAACTTTTACTGCTGCGCCAGAGCATGGGTTTCTTGAGAAACAGCGCCCTCT 666  
Qy 6615 AATTCACTTTGTGAAGTTGGTTTCGAGCAAGCAGGGGCACATGACACCACTATGATTGT 6674  
Db 667 AATTCACTTTGTGAAGTTGGTTTCGAGCAAGCAGGGGCACATGACACCACTATGATTGT 726  
Qy 6675 TCACCTGAGTGTGGAGTTGGAGAACTGAGGATTTTATTTGCTGTGGACCACTTAACACA 6734  
Db 727 TCACCTGCA----- 734  
Qy 6735 ACATATAAATGACCATGATTTTGTGGATATATATGAGCTAGTAGCTGAAGTGA 6794  
Db 735 ----- 734



Db 891 TAATGTAATATTCAGACCAAGGATACATTTGGAAGAGATTTTAAATCCAGGGGCCAA 950  
QY 7035 AGTTACCCCTCATTTCTTCGAAATGTAATGTCACACCTTTAAAGAAATATCTATGCTTCT 7094  
Db 951 AGTTACCCCTCATTTCTTCGAAATGTAATGTCACACCTTTAAAGAAATATCTATGCTTCT 1010  
QY 7095 CTCAGTGTGCTTT 7108  
Db 1011 CTCAGTGTGCTTT 1024

## RESULT 13

US-09-361-096A-6  
; Sequence 6, Application US/09361096A  
; Patent No. 6492495  
; GENERAL INFORMATION:  
; APPLICANT: MOLLER, NIELS P.H.  
; APPLICANT: MOLLER, KARIN B.  
; APPLICANT: ULLRICH, AXEL  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE  
; FILE REFERENCE: 038602/0686  
; CURRENT APPLICATION NUMBER: US/09/361,096A  
; CURRENT FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: 08/449,609  
; PRIOR FILING DATE: 1995-05-24  
; PRIOR APPLICATION NUMBER: 08/036,210  
; PRIOR FILING DATE: 1995-03-23  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2173  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PTP-S31C  
; NAME/KEY: modified base  
; LOCATION: (1)..(1630)  
; OTHER INFORMATION: n = unknown nucleotide  
; NAME/KEY: CDS  
; LOCATION: (65)..(1066)  
US-09-361-096A-6

Query Match 11.1%; Score 789.6; DB 4; Length 2173;  
Best Local Similarity 86.9%; Pred. No. 9.8e-201;  
Matches 951; Conservative 0; Mismatches 4; Indels 139; Gaps 2;  
QY 6015 AACGGGCCAATAAGCAAGAAATCCTTCTCGAACATGTTGAAGAGTTTGCACAAACAA 6074  
Db 70 AATGAGGCCAATAAGCAAGAAATCCTTCTCGAACATGTTGAAGAGTTTGCACAAACAA 129  
QY 6075 CAACCTAAAGTTTCAAGAGAAATTTTCGGAATACCAAAATTTCTTCAGATCTTTCTTC 6134  
Db 130 CAACCTAAAGTTTCAAGAGAAATTTTCGGAATACCAAAATTTCTTCAGATCTTTCTTC 189  
QY 6135 AACTGATGCTGATCTGCTTGGAAATAGACAAACAAACCGTTTCCCAACATAAAACATA 6194  
Db 190 AACTGATGCTGATCTGCTTGGAAATAGACAAACAAACCGTTTCCCAACATAAAACATA 248  
QY 6195 TAAATAATAAACAAGAGTAAGCTGATGATGCTGACGCTAGTGTTCAGAGTTTCGGATTATAT 6254  
Db 249 --ATAATAAACAAGAGTAAGCTGATGATGCTGACGCTAGTGTTCAGAGTTTCGGATTATAT 306  
QY 6255 TAAATGACGCTATATTTCTGTTTATTTATGTCAAATGAATTTATTTGCTACTCAAGTCC 6314  
Db 307 TAAATGACGCTATATTTCTGTTTATTTATGTCAAATGAATTTATTTGCTACTCAAGTCC 366  
QY 6315 ACTACACAGGAACAGTTGGAGATTTTGGGAATGTTGGGAACCGAGCAACAGGCAAAACATT 6374  
Db 367 ACTACACAGGAACAGTTGGAGATTTTGGGAATGTTGGGAACCGAGCAACAGGCAAAACATT 426  
QY 6375 AGTAATGCTAACACAGTGTTTTGAAGAAAGGACCGATCAGATGCGATCAGTATTGGCCAGA 6434

Db 427 AGTAATGCTAACACACAGTGTTTTGAAGAAAGGACGGATCAGATGCGATCAGTATTGGCCAGA 486  
QY 6435 GGAACAACAAGCCAGTTACTGTCTTTTGGAGATATAGTATTAACAAGCTAATGAGGATGT 6494  
Db 487 GGAACAACAAGCCAGTTACTGTCTTTTGGAGATATAGTATTAACAAGCTAATGAGGATGT 546  
QY 6495 TCAATAGATTGAGACTATCAGGGATCTGAAAATTTGAAAGCATGGGATTTGCATGACTGT 6554  
Db 547 TCAATAGATTGAGACTATCAGGGATCTGAAAATTTGAAAGCATGGGATTTGCATGACTGT 606  
QY 6555 TCGACAGGTGTAACCTTTTACTGCTGCGCAGAGCATGGGTTCTTGAGAAACAGCGCCCTCT 6614  
Db 607 TCGACAGGTGTAACCTTTTACTGCTGCGCAGAGCATGGGTTCTTGAGAAACAGCGCCCTCT 666  
QY 6615 AATTCACTTTTGTGAAGTTGGTTTCGACAAAGCAGGCGACATGACACACCACTATGATTGT 6674  
Db 667 AATTCACTTTTGTGAAGTTGGTTTCGACAAAGCAGGCGACATGACACACCACTATGATTGT 726  
QY 6675 TCACTGAGTGTGGAGTTGGAAAGAACTGGAGTTTATTGCTCTGGACCAATTAACACA 6734  
Db 727 TCACTGCA----- 734  
QY 6735 ACATATAAATGACCATGATTTTGTGGATATATATGAGTAGTAGTGAAGTGA 6794  
Db 735 ----- 734  
QY 6795 AAGAATGTGATGGTGCAGAAATCTGGCAGACAGTATATCTTTTACACAGTGCATTTCTGGA 6854  
Db 735 -----GGCAGACAGTATCTTTTACACAGTGCATTTCTGGA 770  
QY 6855 TCTTTATCAAAATAAGGGAAGTAATCAGCCCATCTGTTTGTGTTAACTATTTCAGACATTCA 6914  
Db 771 TCTTTATCAAAATAAGGGAAGTAATCAGCCCATCTGTTTGTGTTAACTATTTCAGACATTCA 830  
QY 6915 GAAGATGGACTCTTTGAGCGCCATGCAAGTGTGATGTTGAGCTTGAATGGGAAGAACCCAC 6974  
Db 831 GAAGATGGACTCTTTGAGCGCCATGCAAGTGTGATGTTGAGCTTGAATGGGAAGAACCCAC 890  
QY 6975 TATGTAATATTTCAGACCAAGGATACAAATTTGGAAGAGATTTTAAATCCAGGGGCCAA 7034  
Db 891 TATGTAATATTTCAGACCAAGGATACAAATTTGGAAGAGATTTTAAATCCAGGGGCCAA 950  
QY 7035 AGTTACCCCTCATTTCTCGAAATGTAATGTCACACCTTTAAAGAAATATCTATGCTTCT 7094  
Db 951 AGTTACCCCTCATTTCTCGAAATGTAATGTCACACCTTTAAAGAAATATCTATGCTTCT 1010  
QY 7095 CTCAGTGTGCTTT 7108  
Db 1011 CTCAGTGTGCTTT 1024

## RESULT 14

US-08-036-210-8  
; Sequence 8, Application US/08036210  
; Patent No. 5585233  
; GENERAL INFORMATION:  
; APPLICANT: Moller, Niels P.H.  
; APPLICANT: Moller, Karin B.  
; APPLICANT: Ullrich, Axel  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASE  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/036,210  
FILING DATE: 23-MAR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7683-025  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 532 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cdna  
US-08-036-210-8

Query Match 7.5%; Score 532; DB 1; Length 532;  
Best Local Similarity 100.0%; Pred. No. 4.6e-132;  
Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 6437 ACAACAGCCAGTACTGCTTTGGAGATATAGTGAATCAAGCTTAATGGAGGATGTTTC 6496
DB 1 ACAACAGCCAGTACTGCTTTGGAGATATAGTGAATCAAGCTTAATGGAGGATGTTTC 60
QY 6497 AAATAGATTGGACTATCAGGGATCTGAAAATTGAAAGGCATGGGATTCGATGACTGTTTC 6556
DB 61 AAATAGATTGGACTATCAGGGATCTGAAAATTGAAAGGCATGGGATTCGATGACTGTTTC 120
QY 6557 GACAGTGTAACCTTTACTGCTGCGCAGAGCATGGGATTTTATGCTCTGGACCACTTTAACACAA 6616
DB 121 GACAGTGTAACCTTTACTGCTGCGCAGAGCATGGGATTTTATGCTCTGGACCACTTTAACACAA 6616
QY 6617 TTCACTTTGTAAGTTGGTTCGAGCAAGCATGAGGATCTGTTTATGCTCTGGACCACTTTAACACAA 6736
DB 181 TTCACTTTGTAAGTTGGTTCGAGCAAGCATGAGGATTTTATGCTCTGGACCACTTTAACACAA 6736
QY 6737 ATATAAATGACCAATGATTTTGGATATATATGGAATGAGGATGAGTGAAGTGAAGTGAAGTGA 6968
DB 301 ATATAAATGACCAATGATTTTGGATATATATGGAATGAGGATGAGTGAAGTGAAGTGAAGTGA 480
QY 6917 AGATGGACTCTTTGGACGCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGA 6968
DB 481 AGATGGACTCTTTGGACGCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGA 532
```

## RESULT 15

US-08-449-609-8  
Sequence 8, Application US/08449609  
Patent No. 5952212  
GENERAL INFORMATION:  
APPLICANT: Moller, Niels P.H.  
APPLICANT: Moller, Karin B.  
APPLICANT: Ullrich, Axel  
TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
TITLE OF INVENTION: PHOSPHATASE

NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449,609  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/036,210  
FILING DATE: 23-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7683-025  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 532 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cdna  
US-08-449-609-8

Query Match 7.5%; Score 532; DB 2; Length 532;

Best Local Similarity 100.0%; Pred. No. 4.6e-132; Mismatches 0; Indels 0; Gaps 0;  
Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 6437 ACAACAGCCAGTACTGCTTTGGAGATATAGTGAATCAAGCTTAATGGAGGATGTTTC 6496
DB 1 ACAACAGCCAGTACTGCTTTGGAGATATAGTGAATCAAGCTTAATGGAGGATGTTTC 60
QY 6497 AAATAGATTGGACTATCAGGGATCTGAAAATTGAAAGGCATGGGATTCGATGACTGTTTC 6556
DB 61 AAATAGATTGGACTATCAGGGATCTGAAAATTGAAAGGCATGGGATTCGATGACTGTTTC 120
QY 6557 GACAGTGTAACCTTTACTGCTGCGCAGAGCATGGGATTTTATGCTCTGGACCACTTTAACACAA 6616
DB 121 GACAGTGTAACCTTTACTGCTGCGCAGAGCATGGGATTTTATGCTCTGGACCACTTTAACACAA 6616
QY 6617 TTCACTTTGTAAGTTGGTTCGAGCAAGCATGAGGATCTGTTTATGCTCTGGACCACTTTAACACAA 6736
DB 181 TTCACTTTGTAAGTTGGTTCGAGCAAGCATGAGGATTTTATGCTCTGGACCACTTTAACACAA 6736
QY 6677 ACTGAGTGTCTGGAGTTGGAGAACTGGAGTTTATGCTCTGGACCACTTTAACACAA 6736
DB 241 ACTGAGTGTCTGGAGTTGGAGAACTGGAGTTTATGCTCTGGACCACTTTAACACAA 300
QY 6737 ATATAAATGACCAATGATTTTGGATATATATGGAATGAGGATGAGTGAAGTGAAGTGAAGTGA 6796
DB 301 ATATAAATGACCAATGATTTTGGATATATATGGAATGAGGATGAGTGAAGTGAAGTGAAGTGA 360
QY 6797 GAATGTGATGTCAGAAATCTGGACAGTATATCTTTTACACAGTGCATTTCTGGATC 6856
DB 361 GAATGTGATGTCAGAAATCTGGACAGTATATCTTTTACACAGTGCATTTCTGGATC 420
QY 6857 TCTTATCAATAAGGAGTGAATCAAGCCATCTGTTTGTGTTTAACTATTTCAGCACTTCAGA 6916
DB 421 TCTTATCAATAAGGAGTGAATCAAGCCATCTGTTTGTGTTTAACTATTTCAGCACTTCAGA 480
```

Qy 6917 AGATGGACTCTTTGGAGCCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGA 6968  
 Db 481 AGATGGACTCTTTGGAGCCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGA 532

Search completed: July 2, 2005, 21:39:34  
Job time : 1621 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 12:07:45 ; Search time 5469 Seconds  
(without alignments)  
8151.890 Million cell updates/sec

Title: US-10-673-885-1

Perfect score: 7108

Sequence: 1 taattgtacttgcagaa.....gcttctcactgtgccttt 7108

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 6313374 seqs, 3136092125 residues

Total number of hits satisfying chosen parameters: 12626748

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/US06\_PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq.\*
- 20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq.\*
- 21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq.\*
- 22: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 23: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq.\*
- 24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Match | Length | ID | Description      |
|------------|--------|-------|--------|----|------------------|
| 1          | 7108   | 100.0 | 7108   | 10 | US-09-822-871-1  |
| 2          | 7108   | 100.0 | 7108   | 18 | US-10-673-885-1  |
| 3          | 6799.8 | 95.7  | 7225   | 18 | US-10-466-759-10 |
| 4          | 6683.6 | 94.0  | 6903   | 17 | US-10-087-684-9  |
| 5          | 6683.6 | 94.0  | 6903   | 17 | US-10-218-779-9  |
| 6          | 6453   | 90.8  | 6994   | 17 | US-10-087-684-5  |
| 7          | 6453   | 90.8  | 6994   | 17 | US-10-218-779-5  |

|    |        |      |        |    |                    |                    |
|----|--------|------|--------|----|--------------------|--------------------|
| 8  | 2618.6 | 36.8 | 3973   | 15 | US-10-314-232-21   | Sequence 21, Appli |
| 9  | 2551   | 35.9 | 2565   | 17 | US-10-087-684-7    | Sequence 7, Appli  |
| 10 | 2551   | 35.9 | 2565   | 17 | US-10-218-779-7    | Sequence 7, Appli  |
| 11 | 1118.8 | 15.7 | 2692   | 15 | US-10-314-232-14   | Sequence 14, Appli |
| 12 | 1071.6 | 15.1 | 2309   | 15 | US-10-314-232-10   | Sequence 10, Appli |
| 13 | 789.6  | 11.1 | 2173   | 15 | US-10-314-232-6    | Sequence 6, Appli  |
| 14 | 532    | 7.5  | 532    | 15 | US-10-314-232-8    | Sequence 8, Appli  |
| 15 | 319.4  | 4.5  | 321    | 15 | US-10-314-232-3    | Sequence 3, Appli  |
| 16 | 315    | 4.4  | 254366 | 10 | US-09-822-871-3    | Sequence 3, Appli  |
| 17 | 315    | 4.4  | 254366 | 18 | US-10-673-885-3    | Sequence 3, Appli  |
| 18 | 193.8  | 2.7  | 6329   | 21 | US-10-497-692-3    | Sequence 3, Appli  |
| 19 | 192.2  | 2.7  | 936    | 18 | US-10-634-027-5    | Sequence 5, Appli  |
| 20 | 192.2  | 2.7  | 1011   | 18 | US-10-634-027-3    | Sequence 3, Appli  |
| 21 | 192.2  | 2.7  | 5069   | 17 | US-10-331-496A-74  | Sequence 74, Appli |
| 22 | 192.2  | 2.7  | 5415   | 19 | US-10-775-169-153  | Sequence 153, App  |
| 23 | 192.2  | 2.7  | 5415   | 19 | US-10-789-241-33   | Sequence 33, Appli |
| 24 | 192.2  | 2.7  | 6075   | 18 | US-10-641-643-1366 | Sequence 1366, Ap  |
| 25 | 192.2  | 2.7  | 6075   | 18 | US-10-634-027-1    | Sequence 1, Appli  |
| 26 | 192.2  | 2.7  | 6331   | 17 | US-10-159-563-238  | Sequence 238, App  |
| 27 | 192.2  | 2.7  | 7444   | 10 | US-09-909-567B-37  | Sequence 37, Appli |
| 28 | 192.2  | 2.7  | 7444   | 16 | US-10-252-157-477  | Sequence 477, App  |
| 29 | 192.2  | 2.7  | 7879   | 14 | US-10-198-846-9851 | Sequence 9851, Ap  |
| 30 | 180.8  | 2.5  | 4871   | 21 | US-10-486-706-242  | Sequence 242, App  |
| 31 | 176.2  | 2.5  | 6314   | 17 | US-10-366-547-48   | Sequence 48, Appli |
| 32 | 171.8  | 2.4  | 4756   | 17 | US-10-366-547-43   | Sequence 43, Appli |
| 33 | 171.8  | 2.4  | 4756   | 18 | US-10-641-643-1461 | Sequence 1461, Ap  |
| 34 | 171.8  | 2.4  | 5117   | 15 | US-10-390-501-1    | Sequence 1, Appli  |
| 35 | 171.8  | 2.4  | 5117   | 17 | US-10-366-547-41   | Sequence 41, Appli |
| 36 | 171.8  | 2.4  | 5117   | 19 | US-10-723-606-1    | Sequence 1, Appli  |
| 37 | 169.2  | 2.4  | 1752   | 15 | US-10-005-220-1    | Sequence 1, Appli  |
| 38 | 167    | 2.3  | 4078   | 17 | US-10-305-720-1134 | Sequence 1134, Ap  |
| 39 | 167    | 2.3  | 4323   | 11 | US-09-997-722-48   | Sequence 48, Appli |
| 40 | 167    | 2.3  | 4651   | 9  | US-09-887-669-4    | Sequence 4, Appli  |
| 41 | 167    | 2.3  | 4651   | 18 | US-10-641-643-945  | Sequence 945, App  |
| 42 | 167    | 2.3  | 5775   | 11 | US-09-997-722-47   | Sequence 47, Appli |
| 43 | 167    | 2.3  | 5775   | 17 | US-10-189-429-4    | Sequence 4, Appli  |
| 44 | 165.4  | 2.3  | 4426   | 17 | US-10-366-547-46   | Sequence 46, Appli |
| 45 | 156.4  | 2.2  | 6115   | 20 | US-10-723-860-5538 | Sequence 5538, Ap  |

ALIGNMENTS

RESULT 1  
US-09-822-871-1  
; Sequence 1, Application US/09822871  
; Publication No. US20030099942A1  
; GENERAL INFORMATION:  
; APPLICANT: WEBSTER, Marion et al  
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CLO01219  
; CURRENT APPLICATION NUMBER: US/09/822,871  
; CURRENT FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 7108  
; TYPE: DNA  
; ORGANISM: Human  
US-09-822-871-1

Query Match 100.0%; Score 7108; DB 10; Length 7108;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 7108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TAATTGTGCTACTTGCAGAGATCTGCTTTAAATCAATTAATGAGCAACATTCTCT 60  
DB 1 TAATTGTGCTACTTGCAGAGATCTGCTTTAAATCAATTAATGAGCAACATTCTCT 60  
QY 61 CTAGAGCCATCAATGTGATTTCTACTGGTGAATAAATGTAATAAGATGATTTCTTATC 120



Db 61 CTAGAGCCATCAATGTGATTTCTACTGGCTGAAAAATGTAAATAAGATGGATTTTCTTATC 120  
Qy 121 ATTTTCTTTTATCTTTTATTTGGACTTCAGAGACACAGGTTGATGTTTCCAAATGTCGTT 180  
Db 121 ATTTTCTTTTATCTTTTATTTGGACTTCAGAGACACAGGTTGATGTTTCCAAATGTCGTT 180  
Qy 181 CTTGGTACTAGGTACGATATAACCATCTCTTCAATTTCTACAACATACACCTCACCTGTT 240  
Db 181 CTTGGTACTAGGTACGATATAACCATCTCTTCAATTTCTACAACATACACCTCACCTGTT 240  
Qy 241 ACTAGAAATAGTGACCAAAATGTAAACAAACAGGGCTTCCAGTCTTCTTACCGGGGAA 300  
Db 241 ACTAGAAATAGTGACCAAAATGTAAACAAACAGGGCTTCCAGTCTTCTTACCGGGGAA 300  
Qy 301 AGAGTCGGATCTGCTGGGATCTTCTGCTTGGATATACACACCTAAATCCAAATGGAGG 360  
Db 301 AGAGTCGGATCTGCTGGGATCTTCTGCTTGGATATACACACCTAAATCCAAATGGAGG 360  
Qy 361 ATTATATCTTACATTTGCTCAAAATATAAGGAAGTTTGTCCGTGGATGCAACAGTATATACA 420  
Db 361 ATTATATCTTACATTTGCTCAAAATATAAGGAAGTTTGTCCGTGGATGCAACAGTATATACA 420  
Qy 421 CAAGTCAGATCAAGCCAGACAGTCTGGAAGTTCTTCTTAACTTAACTCTTAACTCTGGAACA 480  
Db 421 CAAGTCAGATCAAGCCAGACAGTCTGGAAGTTCTTCTTAACTTAACTCTTAACTCTGGAACA 480  
Qy 481 ACATATGAATTAAGGTTCTGCTGAAACAGTCTGGCAATGGAGTGTATAGTATCCA 540  
Db 481 ACATATGAATTAAGGTTCTGCTGAAACAGTCTGGCAATGGAGTGTATAGTATCCA 540  
Qy 541 TTTCTCTTCCAAACTCGAAGAGTCTCCAGGAAAGTGGTGAATCTCAAGTTGAGGCC 600  
Db 541 TTTCTCTTCCAAACTCGAAGAGTCTCCAGGAAAGTGGTGAATCTCAAGTTGAGGCC 600  
Qy 601 TACAACGCTTCAGCAGTTAAGCTGATTTGGTATTTTACCTCGCAACCAAAATGGCAAAAT 660  
Db 601 TACAACGCTTCAGCAGTTAAGCTGATTTGGTATTTTACCTCGCAACCAAAATGGCAAAAT 660  
Qy 661 ACCAGCTTCAAGTTAGTGTCAACATGCGAGAGTGGATGAGTAGTAGTAAGATGTCTCA 720  
Db 661 ACCAGCTTCAAGTTAGTGTCAACATGCGAGAGTGGATGAGTAGTAGTAAGATGTCTCA 720  
Qy 721 ATCAGAGTAGAGACATTTGACTGGGAAATGGCAGAGATGCAATGAGAAATAGTAATCT 780  
Db 721 ATCAGAGTAGAGACATTTGACTGGGAAATGGCAGAGATGCAATGAGAAATAGTAATCT 780  
Qy 781 TTTTATGAGGATACAGCCAGCCCTTCTCAAACCTTTGGTGGATGACCTCCATCGCGT 840  
Db 781 TTTTATGAGGATACAGCCAGCCCTTCTCAAACCTTTGGTGGATGACCTCCATCGCGT 840  
Qy 841 ACCACATTTCAAGCAGCTGTGACACAGATGAGATCAGCTCTGTGTGGAAAGGCT 900  
Db 841 ACCACATTTCAAGCAGCTGTGACACAGATGAGATCAGCTCTGTGTGGAAAGGCT 900  
Qy 901 ATCAGTTTGTAGTGACACATTTGAGACCTTATACAAATATCTTTTGAAGTTTCAGCT 960  
Db 901 ATCAGTTTGTAGTGACACATTTGAGACCTTATACAAATATCTTTTGAAGTTTCAGCT 960  
Qy 961 GCTAACAATGAAAGAGGTTTATTTGATAGTACGATTTGTCAGAAACACCAAGATCAGTGCCT 1020  
Db 961 GCTAACAATGAAAGAGGTTTATTTGATAGTACGATTTGTCAGAAACACCAAGATCAGTGCCT 1020  
Qy 1021 GAAGNCCACCAAACTGCGTTACAGGCAACATACAGGAAAGTCCCTTTCAATTTTA 1080  
Db 1021 GAAGNCCACCAAACTGCGTTACAGGCAACATACAGGAAAGTCCCTTTCAATTTTA 1080  
Qy 1081 TGGGACCCACCAACTATAGTAAACAGGAAATTTAGTTATAGAGTTGAAATATATGAGCA 1140  
Db 1081 TGGGACCCACCAACTATAGTAAACAGGAAATTTAGTTATAGAGTTGAAATATATGAGCA 1140  
Qy 1141 TCAGTTCGATTTTGGATTAACGCAAAAGACCTCAAGTTTGGATTTCACTAACCTTAACA 1200  
Db 1141 TCAGTTCGATTTTGGATTAACGCAAAAGACCTCAAGTTTGGATTTCACTAACCTTAACA 1200

Qy 1201 CCATTTTCAATCTATGATGTTCTATATTCGGCTGAAACCCAGTGCAGGACTCGGCCCAAG 1260  
Db 1201 CCATTTTCAATCTATGATGTTCTATATTCGGCTGAAACCCAGTGCAGGACTCGGCCCAAG 1260  
Qy 1261 TCAAAATATTTTCAGTATTCACCTCCACCAGATGTTCCAGGGGAGTGTGTTGATTTACAACCT 1320  
Db 1261 TCAAAATATTTTCAGTATTCACCTCCACCAGATGTTCCAGGGGAGTGTGTTGATTTACAACCT 1320  
Qy 1321 GCAGAGGTAGAAATCCACGCAAGTAAGAATTTCTTGGAAAGAAACACGACCAACAAATGGA 1380  
Db 1321 GCAGAGGTAGAAATCCACGCAAGTAAGAATTTCTTGGAAAGAAACACGACCAACAAATGGA 1380  
Qy 1381 ATTATTAACCAATACCGAGTGAAGTGTAGTTCCAGAGACAGGAATTAATTTTGGAAAT 1440  
Db 1381 ATTATTAACCAATACCGAGTGAAGTGTAGTTCCAGAGACAGGAATTAATTTTGGAAAT 1440  
Qy 1441 ACTTTGCTCCTCTGGAATTAATAGATATATAAATGACCCCATGGCTCCAGAAATTTGTGAAC 1500  
Db 1441 ACTTTGCTCCTCTGGAATTAATAGATATATAAATGACCCCATGGCTCCAGAAATTTGTGAAC 1500  
Qy 1501 ATAGTAGAGCCAAATGCTAGGATTTATGAGGGTTCCAGAGAGATGTCGTGACCTTCAC 1560  
Db 1501 ATAGTAGAGCCAAATGCTAGGATTTATGAGGGTTCCAGAGAGATGTCGTGACCTTCAC 1560  
Qy 1561 TCACCTGCTACATTTTATATAAAGCCATCCAGATAAATACTTTCTCGAAGGAATAGA 1620  
Db 1561 TCACCTGCTACATTTTATATAAAGCCATCCAGATAAATACTTTCTCGAAGGAATAGA 1620  
Qy 1621 GCTGAAGACCCAGACTTCACAGTTGTAATACAAGAAATCAGTATATTTACTGACATTTGCA 1680  
Db 1621 GCTGAAGACCCAGACTTCACAGTTGTAATACAAGAAATCAGTATATTTACTGACATTTGCA 1680  
Qy 1681 GCTGAACAGCTGCTTTATGTTATCAGGAGACTTTGTAACCTTTTCACTGAGCACAATGATAGT 1740  
Db 1681 GCTGAACAGCTGCTTTATGTTATCAGGAGACTTTGTAACCTTTTCACTGAGCACAATGATAGT 1740  
Qy 1741 GTATCTGCTTTTACCACATCATGGAGAGGACCAACAGTTCTCAGTGTGTAGGACAGT 1800  
Db 1741 GTATCTGCTTTTACCACATCATGGAGAGGACCAACAGTTCTCAGTGTGTAGGACAGT 1800  
Qy 1801 CAGCAAGTGCACAGCTCCATTTAAATATAAATATAAATAATTTAGTTCTTCATCTATT 1860  
Db 1801 CAGCAAGTGCACAGCTCCATTTAAATATAAATATAAATAATTTAGTTCTTCATCTATT 1860  
Qy 1861 TTGTTATATTTGGGATCTCCAGATATCCCAATGGAATAATTAACCTACTATACGATTTAT 1920  
Db 1861 TTGTTATATTTGGGATCTCCAGATATCCCAATGGAATAATTAACCTACTATACGATTTAT 1920  
Qy 1921 GCAATGGAATTTGGATACAAACAGACTTTCCAGATAAATACCATAAGATAAAGCTTTCTC 1980  
Db 1921 GCAATGGAATTTGGATACAAACAGACTTTCCAGATAAATACCATAAGATAAAGCTTTCTC 1980  
Qy 1981 ATAAACAGGGTTAAAGAAATACAAAATACAAAATGAGAGTGGAGCTCAACCCACGAT 2040  
Db 1981 ATAAACAGGGTTAAAGAAATACAAAATACAAAATGAGAGTGGAGCTCAACCCACGAT 2040  
Qy 2041 GGAGAAAGTTCTTTGCTGGAAGAAATGACATCTTTTGTGAGAACTTCAGAGATGAACCG 2100  
Db 2041 GGAGAAAGTTCTTTGCTGGAAGAAATGACATCTTTTGTGAGAACTTCAGAGATGAACCG 2100  
Qy 2101 GAATCATCACTCAAGATGTCGAAGTAATGATGTTTACCAGAGATGAATAAGGTTTGAAG 2160  
Db 2101 GAATCATCACTCAAGATGTCGAAGTAATGATGTTTACCAGAGATGAATAAGGTTTGAAG 2160  
Qy 2161 TGGTCAACCCGAAAGCCCAATGCGGATCAATTTTGTCTTATGAAGTGTCTATATAAAT 2220  
Db 2161 TGGTCAACCCGAAAGCCCAATGCGGATCAATTTTGTCTTATGAAGTGTCTATATAAAT 2220  
Qy 2221 ATAGATATCTTTTATATATGAAGAAACACATCAACAGACATATATTTAAGGAACTTAAGA 2280  
Db 2221 ATAGATATCTTTTATATATGAAGAAACACATCAACAGACATATATTTAAGGAACTTAAGA 2280





```
QY 6661 ACACCTATGATGTTGCTGACGTGCTGGAGTGGAGAACTGGAGTGTATGCTCTG 6720
Db 6661 ACACCTATGATGTTGCTGACGTGCTGGAGTGGAGAACTGGAGTGTATGCTCTG 6720
QY 6721 GACCATTTAAACACACATATAAATGACCATGATTTTGTGGATATATATGAGCTAGTAGCT 6780
Db 6721 GACCATTTAAACACACATATAAATGACCATGATTTTGTGGATATATATGAGCTAGTAGCT 6780
QY 6781 GAACTGAGAAGTGAAGAATGTCATGGTGCAGAACTGGCACAGTATATCTTTTACAC 6840
Db 6781 GAACTGAGAAGTGAAGAATGTCATGGTGCAGAACTGGCACAGTATATCTTTTACAC 6840
QY 6841 CAGTGATCTGATCTCTATCAATTAAGGAGTAATCAGGCCATCTGTTTGTAAAC 6900
Db 6841 CAGTGATCTGATCTCTATCAATTAAGGAGTAATCAGGCCATCTGTTTGTAAAC 6900
QY 6901 TATTGAGCACTTCAGAAAGTGAAGTCTTTGGAGCCATGGAAGTGTGAGCTTGA 6960
Db 6901 TATTGAGCACTTCAGAAAGTGAAGTCTTTGGAGCCATGGAAGTGTGAGCTTGA 6960
QY 6961 TGGGAAGAAACCACTATGTAATATTTCAGACCAAGGATACAAATGGAAAGATTTTAA 7020
Db 6961 TGGGAAGAAACCACTATGTAATATTTCAGACCAAGGATACAAATGGAAAGATTTTAA 7020
QY 7021 ATCCAGGGGCCAAAGTTACCCCTCATCTCCGAATGGAATGTGCAACCTTAAAGAA 7080
Db 7021 ATCCAGGGGCCAAAGTTACCCCTCATCTCCGAATGGAATGTGCAACCTTAAAGAA 7080
QY 7081 ATATCTATGCTCTCTCACTGTCCTTT 7108
Db 7081 ATATCTATGCTCTCTCACTGTCCTTT 7108
```

## RESULT 2

```
US-10-673-885-1
; Sequence 1, Application US/10673885
; Publication No. US20040081644A1
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001219DIV
; CURRENT APPLICATION NUMBER: US/10/673,885
; CURRENT FILING DATE: 2003-09-30
; PRIOR APPLICATION NUMBER: 09/822,871
; PRIOR FILING DATE: 04-02-2001
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 7108
; TYPE: DNA
; ORGANISM: Human
US-10-673-885-1
```

```
Query Match 100.0%; Score 7108; DB 18; Length 7108;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 7108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAATTGTGACTTGCAGAAAGATCTGCTTTAAATCAATTAATGCAAGGCAACATTTCTCT 60
Db 1 TAATTGTGACTTGCAGAAAGATCTGCTTTAAATCAATTAATGCAAGGCAACATTTCTCT 60
QY 61 CTAGAGCCATCAATGATGATCTGCTGGAAGTGAATTAAGATGATTTCTTATC 120
Db 61 CTAGAGCCATCAATGATGATCTGCTGGAAGTGAATTAAGATGATTTCTTATC 120
QY 121 ATTTTCTTTTACTTTTATTTGGACTTCAGAGACACAGGTTGATGTTTCCAATGTCGTT 180
Db 121 ATTTTCTTTTACTTTTATTTGGACTTCAGAGACACAGGTTGATGTTTCCAATGTCGTT 180
QY 181 CCTGGTACTAGGTACGATATAACCATCTCTTCAATTTCTTCAACATACACCTCACTGTT 240
Db 181 CCTGGTACTAGGTACGATATAACCATCTCTTCAATTTCTTCAACATACACCTCACTGTT 240
```

```
Db 181 CCTGGTACTAGGTACGATATAACCATCTCTTCAATTTCTTCAACATACACCTCACTGTT 240
QY 241 ACTAGATAGTGCACCAAAATGTAACAAACAGAGGCTCTCCAGTCTTCTTAGCCGGGAA 300
Db 241 ACTAGATAGTGCACCAAAATGTAACAAACAGAGGCTCTCCAGTCTTCTTAGCCGGGAA 300
QY 301 AGAGTCGGATCTGCTGGGATTTCTTCTGCTTGGATACACCACTTAATCAATGGAAGG 360
Db 301 AGAGTCGGATCTGCTGGGATTTCTTCTGCTTGGATACACCACTTAATCAATGGAAGG 360
QY 361 ATTATATCTTACATTTGTCAAATATTAAGGAAGTTTGTCCGTGATGCAACAGTATATACA 420
Db 361 ATTATATCTTACATTTGTCAAATATTAAGGAAGTTTGTCCGTGATGCAACAGTATATACA 420
QY 421 CAAAGTCAGATCAAAAGCCAGACAGTCTGGAAGTTCTTTTAACTTAATCTTAACTCTGGAACA 480
Db 421 CAAAGTCAGATCAAAAGCCAGACAGTCTGGAAGTTCTTTTAACTTAATCTTAACTCTGGAACA 480
QY 481 ACATATGAAATTAAGGTTGCTGCTGAAACAGTCTGGAATTTGGATTTGGATTTAGTATCCA 540
Db 481 ACATATGAAATTAAGGTTGCTGCTGAAACAGTCTGGAATTTGGATTTGGATTTAGTATCCA 540
QY 541 TTTCTCTTCCAACTGCAGAAAGTCTCCAGGAAAGTGTGGAATCTCACAGTTGAGGCC 600
Db 541 TTTCTCTTCCAACTGCAGAAAGTGTGGAATCTCCAGGAAAGTGTGGAATCTCACAGTTGAGGCC 600
QY 601 TACAACGCTTCAGCAGTTAAGCTGATTTGGTATTTTACCTCGGCAACCAATGGCAAAAT 660
Db 601 TACAACGCTTCAGCAGTTAAGCTGATTTGGTATTTTACCTCGGCAACCAATGGCAAAAT 660
QY 661 ACCAGCTTCAAGATTTAGTGTCAAAATGTCAGAAAGTGTGGAATCTCACAGTTGAGGCC 720
Db 661 ACCAGCTTCAAGATTTAGTGTCAAAATGTCAGAAAGTGTGGAATCTCACAGTTGAGGCC 720
QY 721 ATCAGATAGGAGCAATTTGATCTGGAAATTTGCCAGATGCAATGAGATAGTGAATCT 780
Db 721 ATCAGATAGGAGCAATTTGATCTGGAAATTTGCCAGATGCAATGAGATAGTGAATCT 780
QY 781 TTTTATGAGTACAGCAGCCCTTCTCCAACTTTGAGTGTAGTGTACACCTCCATCGCT 840
Db 781 TTTTATGAGTACAGCAGCCCTTCTCCAACTTTGAGTGTAGTGTACACCTCCATCGCT 840
QY 841 ACCACATTTCAATCAAGCAAGTGTGACACAGATGAGTCAAGTCTGCTGTGGAAGAGCCT 900
Db 841 ACCACATTTCAATCAAGCAAGTGTGACACAGATGAGTCAAGTCTGCTGTGGAAGAGCCT 900
QY 901 ATCAGTTTGTAGTGCACACATTTGAGCTTATACAAATATCTTTTGAAGTTTCAGCT 960
Db 901 ATCAGTTTGTAGTGCACACATTTGAGCTTATACAAATATCTTTTGAAGTTTCAGCT 960
QY 961 GCTCAACTGAAGCAGGTTTATATGATAGTACGATTTGTGAGAAACCAACAGATCAGTGCCT 1020
Db 961 GCTCAACTGAAGCAGGTTTATATGATAGTACGATTTGTGAGAAACCAACAGATCAGTGCCT 1020
QY 1021 GAAGGACCAACCAAACTGCGTAA CAGGCAACATCAGGAAAGTCTTTTCAATTTTA 1080
Db 1021 GAAGGACCAACCAAACTGCGTAA CAGGCAACATCAGGAAAGTCTTTTCAATTTTA 1080
QY 1081 TGGGACCCCAACCACTATAGTAA CAGGAAATTTAGTATAGATTTGAATTTATGGAACCA 1140
Db 1081 TGGGACCCCAACCACTATAGTAA CAGGAAATTTAGTATAGATTTGAATTTATGGAACCA 1140
QY 1141 TCAGGTGCGATTTTGGATAA CAGCAAAAGACCTCAAGTTTGCATTCTCACTAACCTAAACA 1200
Db 1141 TCAGGTGCGATTTTGGATAA CAGCAAAAGACCTCAAGTTTGCATTCTCACTAACCTAAACA 1200
QY 1201 CCAATTTCAATATGATGTTCTATATTCGGCTGAAACAGTGCAGGAGCTGGGCCCAAG 1260
Db 1201 CCAATTTCAATATGATGTTCTATATTCGGCTGAAACAGTGCAGGAGCTGGGCCCAAG 1260
QY 1261 TCAATATTTTCAATATGATGTTCTATATTCGGCTGAAACAGTGCAGGAGCTGGGCCCAAG 1320
Db 1261 TCAATATTTTCAATATGATGTTCTATATTCGGCTGAAACAGTGCAGGAGCTGGGCCCAAG 1320
```

1321 GCAGAGGTAGATCCACGCAAGTAAGAAATTACTTTGGAAGAAACCCAGCAACCAAAATGGA 1380  
1321 GCAGAGGTAGATCCACGCAAGTAAGAAATTACTTTGGAAGAAACCCAGCAACCAAAATGGA 1380  
1381 ATTATTAAACCAATACCGAGTGAAGTGCTAGTTTCCAGAGACAGGAATAATTTTGGAAAT 1440  
1381 ATTATTAAACCAATACCGAGTGAAGTGCTAGTTTCCAGAGACAGGAATAATTTTGGAAAT 1440  
1441 ACTTTGCTCACTGGAAATTAATGAGTATATTAATGACCCCATGGCTCCAGAAATTTGTGAAC 1500  
1441 ACTTTGCTCACTGGAAATTAATGAGTATATTAATGACCCCATGGCTCCAGAAATTTGTGAAC 1500  
1501 ATAGTAGAGCAATGTTAGGATTATATGAGGTTTACAGAGAGATGTCGTCTGACCTTTCAC 1560  
1501 ATAGTAGAGCAATGTTAGGATTATATGAGGTTTACAGAGAGATGTCGTCTGACCTTTCAC 1560  
1561 TCACTTGTCTACATTTATATATATAACAGCCATCCAGATAAAAACTTTCTGCAAGGAATAGA 1620  
1621 GCTGAAGACAGACTTTCACAGTGTGTAATCTACAGGAATCAGTATATTAATCTGACATTGCA 1680  
1621 GCTGAAGACAGACTTTCACAGTGTGTAATCTACAGGAATCAGTATATTAATCTGACATTGCA 1680  
1681 GCTGAACAGCTGCTTATGTTATCAGGAGACTTGTACCTTTCACTGAGCACAATGATTAGT 1740  
1681 GCTGAACAGCTGCTTATGTTATCAGGAGACTTGTACCTTTCACTGAGCACAATGATTAGT 1740  
1741 GTATCTGCTTTCACCATATGCGGAGAGGACCAACCAAGTCTCTCACTGTTAGGACACGT 1800  
1741 GTATCTGCTTTCACCATATGCGGAGAGGACCAACCAAGTCTCTCACTGTTAGGACACGT 1800  
1801 CAGCAAGTGCAGCTCCATTTAAATTTAACTATAAAATTAATTAATTAATTAATTAATTAAT 1860  
1801 CAGCAAGTGCAGCTCCATTTAAATTTAACTATAAAATTAATTAATTAATTAATTAATTAAT 1860  
1861 TTGTTATATTTGGATCTCCAGAAATCCCAATGGAATAATACTCACTATACGATTTAT 1920  
1861 TTGTTATATTTGGATCTCCAGAAATCCCAATGGAATAATACTCACTATACGATTTAT 1920  
1921 GCAATGGAAATGGATACAAACAGAGCAATCCAGATAAATACCATAGATAACAGCTTCTC 1980  
1921 GCAATGGAAATGGATACAAACAGAGCAATCCAGATAAATACCATAGATAACAGCTTCTC 1980  
1981 ATAAAGGCTTAAAGAAATACAAATAACAAATACAAATACAAATACAAATACAAATACAAAT 2040  
1981 ATAAAGGCTTAAAGAAATACAAATAACAAATACAAATACAAATACAAATACAAATACAAAT 2040  
2041 GGAGAAAGTCTTTGTTGTAAGAAATGACATCTTTGTAAGAAATGACATCTTTGTAAGAAAT 2100  
2041 GGAGAAAGTCTTTGTTGTAAGAAATGACATCTTTGTAAGAAATGACATCTTTGTAAGAAAT 2100  
2101 GAATCATCACTCAAGATGTCGAAGTAAATGATGTTTACCGAGATGAATAAGTTTGAAG 2160  
2101 GAATCATCACTCAAGATGTCGAAGTAAATGATGTTTACCGAGATGAATAAGTTTGAAG 2160  
2161 TGGTCAACCCGAAAGCCCAATGGGATCATTTATGCTTATGCAAGTGCATATAAATAAT 2220  
2161 TGGTCAACCCGAAAGCCCAATGGGATCATTTATGCTTATGCAAGTGCATATAAATAAT 2220  
2221 ATAGATACCTTTATATGTAAGAACACATCAACAGACATATAATTAAGGAACCTTAAGA 2280  
2221 ATAGATACCTTTATATGTAAGAACACATCAACAGACATATAATTAAGGAACCTTTAAGA 2280  
2281 CCTCACACCTCTATAAATTTCTGTAAGGTCTTACACAGATTTGGTCATGGCAATCAG 2340  
2281 CCTCACACCTCTATAAATTTCTGTAAGGTCTTACACAGATTTGGTCATGGCAATCAG 2340  
2341 GTATCTTCTTACTCTGTAGGACTTCGGAGACTGTCGCTGATAGTGCACAGAAAT 2400  
2341 GTATCTTCTTACTCTGTAGGACTTCGGAGACTGTCGCTGATAGTGCACAGAAAT 2400

2401 ATCACTTCAAAAAATATTTCTTCTGGAGAGATTGAGCTATCATTTCTTCCCCCAAGTAGT 2460  
2401 ATCACTTCAAAAAATATTTCTTCTGGAGAGATTGAGCTATCATTTCTTCCCCCAAGTAGT 2460  
2461 CCCAATGGAAATCATAAAAAATATACAATTTATCTCAAGAGAAATGGAATGAGGAA 2520  
2461 CCCAATGGAAATCATAAAAAATATACAATTTATCTCAAGAGAAATGGAATGAGGAA 2520  
2521 AGAATATATAATACAACTCTTTAACCCAAACATTTAAAGTACTGAAGAAATATACCCAA 2580  
2521 AGAATATATAATACAACTCTTTAACCCAAACATTTAAAGTACTGAAGAAATATACCCAA 2580  
2581 TATATCATTTGAGGTCTGCTAGTAGTACACTGAAAGGTGAAGGAGTTGCGAGTCTCCCAT 2640  
2581 TATATCATTTGAGGTCTGCTAGTAGTACACTGAAAGGTGAAGGAGTTGCGAGTCTCCCAT 2640  
2641 AGTATCATGACGAGGAAAGATGCTCTGANTCTCCCTCTCAAGACTTCTCTGTAAAAACAG 2700  
2641 AGTATCATGACGAGGAAAGATGCTCTGANTCTCCCTCTCAAGACTTCTCTGTAAAAACAG 2700  
2701 TTGCTCTGCTGTCACCGTGAAGTTGTCATGGCAACCAACCCCTGGAGCCAAATGGAATATC 2760  
2701 TTGCTCTGCTGTCACCGTGAAGTTGTCATGGCAACCAACCCCTGGAGCCAAATGGAATATC 2760  
2761 CTTTATTTACACAGTTTATGCTCTGGAATAGATCATATTTAAAAAATTAATGCTCACTGAA 2820  
2761 CTTTATTTACACAGTTTATGCTCTGGAATAGATCATATTTAAAAAATTAATGCTCACTGAA 2820  
2821 ACATCATTTGGAGTTATCAGATTGGAATTAATGTTGAAATACAGTGTCTTATGTAACAGCT 2880  
2821 ACATCATTTGGAGTTATCAGATTGGAATTAATGTTGAAATACAGTGTCTTATGTAACAGCT 2880  
2881 AGCACCAAGTTTGGTGAAGGGAACAGGAGCAATATCATTTAGCTTTTCAACACACAG 2940  
2881 AGCACCAAGTTTGGTGAAGGGAACAGGAGCAATATCATTTAGCTTTTCAACACACAG 2940  
2941 GGAGCAACCAAGCGATCTCTCCAAAGATGTTTATTTATGCAAACTCAGTTCCTTCATCAATA 3000  
2941 GGAGCAACCAAGCGATCTCTCCAAAGATGTTTATTTATGCAAACTCAGTTCCTTCATCAATA 3000  
3001 ATTCTTTTCTGGACACCTCTCTTCAAAACCTTAATGGGATTTACAATATTAATCTCTGTTAT 3060  
3001 ATTCTTTTCTGGACACCTCTCTTCAAAACCTTAATGGGATTTACAATATTAATCTCTGTTAT 3060  
3061 TACAGAAATACCTTCAAGTACTTTTATGAGAAATTTTACCTCCATGAACTAACCAATGAC 3120  
3061 TACAGAAATACCTTCAAGTACTTTTATGAGAAATTTTACCTCCATGAACTAACCAATGAC 3120  
3121 TTTGCAATATGACTGTATCCAAATATAGATAAACTGACAATATTCAGCTTACTATACA 3180  
3121 TTTGCAATATGACTGTATCCCAATATAGATAAACTGACAATATTCAGCTTACTATACA 3180  
3181 TTTTGGTTAAACGAAAGTACTTTCAGTTGGAATGGAATTAAGCAGTGCATCAATTGAA 3240  
3181 TTTTGGTTAAACGAAAGTACTTTCAGTTGGAATGGAATTAAGCAGTGCATCAATTGAA 3240  
3241 GTATACACAGATCAAGACATACCTCAAGAGTGTGTTGGAACCTCAGCTTACGAAATCAAT 3300  
3241 GTATACACAGATCAAGACATACCTCAAGAGTGTGTTGGAACCTCAGCTTACGAAATCAAT 3300  
3301 TCGTCAACTGCAATTAATGTAAGCTGGGTCCCAACCGGCTCAACCAACCGTCTAGTCTTC 3360  
3301 TCGTCAACTGCAATTAATGTAAGCTGGGTCCCAACCGGCTCAACCAACCGTCTAGTCTTC 3360  
3361 TACTATGTTTCACTGATCTTACAGCAGACTCTCTGCGCATGTGAGACCACTCTTGTGTA 3420  
3361 TACTATGTTTCACTGATCTTACAGCAGACTCTCTGCGCATGTGAGACCACTCTTGTGTA 3420  
3421 TATGAGAGAGCATATATTTTATGATAATCTGGAATAATACCTGATTTATATTTAAAAAT 3480  
3421 TATGAGAGAGCATATATTTTATGATAATCTGGAATAATACCTGATTTATATTTAAAAAT 3480  
3481 ACTCCATCAACAGAAAAAGGATTTCTTGATACCTTATCTGACCTTATACATCAAGACT 3540





QY 5701 CCAAAAAGCAATCTATTAAATTTAGAGCTACAAAATATTATGGGACAAATTTACTGAC 5760  
DB |||||||  
QY 5701 CCAAAAAGCAATCTATTAAATTTAGAGCTACAAAATATTATGGGACAAATTTACTGAC 5760  
DB |||||||  
QY 5761 TCTGATTTCTGACCCCTGTTAAGACTTTTAGGGAGGAGCTTTTCAGAAAAGAACCGTAGAG 5820  
DB |||||||  
QY 5761 TCTGATTTCTGACCCCTGTTAAGACTTTTAGGGAGGAGCTTTTCAGAAAAGAACCGTAGAG 5820  
DB |||||||  
QY 5821 ATCATCTCTTCGGTCACTTTGTTGATCTCTTCAATTAATTTCTCTTGGACAGCTATTTT 5880  
DB |||||||  
QY 5821 ATCATCTCTTCGGTCACTTTGTTGATCTCTTCAATTAATTTCTCTTGGACAGCTATTTT 5880  
DB |||||||  
QY 5881 GCATTTGCAAGAAATTCGACAGAGCAAGAAAGGTGGCACATCTCTCTCAGGATGCA 5940  
DB |||||||  
QY 5881 GCATTTGCAAGAAATTCGACAGAGCAAGAAAGGTGGCACATCTCTCTCAGGATGCA 5940  
DB |||||||  
QY 5941 GAAATTTATGACACTAAATTTGAAGCTGGATCGCTCATCATCAGTGGCAGACCTGGAAC 6000  
DB |||||||  
QY 5941 GAAATTTATGACACTAAATTTGAAGCTGGATCGCTCATCATCAGTGGCAGACCTGGAAC 6000  
DB |||||||  
QY 6001 AAGGACGAGAGATTACGGGGGCCAATAAGCAAGAAATCCCTCTCTGCAACATGTTGAAGAG 6060  
DB |||||||  
QY 6001 AAGGACGAGAGATTACGGGGGCCAATAAGCAAGAAATCCCTCTCTGCAACATGTTGAAGAG 6060  
DB |||||||  
QY 6061 CTTTGCACAAACAACTAAAGTTTCAAGAGAAATTTTCGGAATTTACCAAAATTTCTT 6120  
DB |||||||  
QY 6061 CTTTGCACAAACAACTAAAGTTTCAAGAGAAATTTTCGGAATTTACCAAAATTTCTT 6120  
DB |||||||  
QY 6121 CAGGATCTTTCTTCAACTGATCTGATCTGCCCTTGGATAGAGCAAAAACCGTTTCCCA 6180  
DB |||||||  
QY 6121 CAGGATCTTTCTTCAACTGATCTGATCTGCCCTTGGATAGAGCAAAAACCGTTTCCCA 6180  
DB |||||||  
QY 6181 AACATTAACCATATATTAATAATACAGAGTAAGCTGATAGCTGACGCTAGTGTCCA 6240  
DB |||||||  
QY 6181 AACATTAACCATATATTAATAATACAGAGTAAGCTGATAGCTGACGCTAGTGTCCA 6240  
DB |||||||  
QY 6241 GGTTCGGATTATTAATTAATCCAGCTATATTTCTCGTTATTTATGTCCTCAATGAATTTAT 6300  
DB |||||||  
QY 6241 GGTTCGGATTATTAATTAATCCAGCTATATTTCTCGTTATTTATGTCCTCAATGAATTTAT 6300  
DB |||||||  
QY 6301 GCTACTCAAGGTCCTACTACAGGAAACAGTTGGAGATTTTGGAGATGGTGTGGGAAACC 6360  
DB |||||||  
QY 6301 GCTACTCAAGGTCCTACTACAGGAAACAGTTGGAGATTTTGGAGATGGTGTGGGAAACC 6360  
DB |||||||  
QY 6361 AGGGCAAAAACATTAGTATGCTAACAAGTGTGTTTTGAAAAAGGACGGATCAGATGCCAT 6420  
DB |||||||  
QY 6361 AGGGCAAAAACATTAGTATGCTAACAAGTGTGTTTTGAAAAAGGACGGATCAGATGCCAT 6420  
DB |||||||  
QY 6421 CAGTATTTGGCCAGAGACAAAGCCAGTTACTGCTTTTGGAGATATAGTGAATTAACAAG 6480  
DB |||||||  
QY 6421 CAGTATTTGGCCAGAGACAAAGCCAGTTACTGCTTTTGGAGATATAGTGAATTAACAAG 6480  
DB |||||||  
QY 6481 CTAATGGAGGATGTTCAAAATAGATTGGACTATCAGGGATCTGAAAATTTGAAAGGCGATGGG 6540  
DB |||||||  
QY 6481 CTAATGGAGGATGTTCAAAATAGATTGGACTATCAGGGATCTGAAAATTTGAAAGGCGATGGG 6540  
DB |||||||  
QY 6541 GATTGATGATGTTTCGACAGTGTAACTTTACTGCTGCGCCAGAGATGGGTTCTGAG 6600  
DB |||||||  
QY 6541 GATTGATGATGTTTCGACAGTGTAACTTTACTGCTGCGCCAGAGATGGGTTCTGAG 6600  
DB |||||||  
QY 6601 AACAGGCCCCCTTAATTTCACTTTGTTGAAGTTGGTTCGAGCAAGCAAGGGCACATGACACC 6660  
DB |||||||  
QY 6601 AACAGGCCCCCTTAATTTCACTTTGTTGAAGTTGGTTCGAGCAAGCAAGGGCACATGACACC 6660  
DB |||||||  
QY 6661 ACACCTATGATGTTTCACTGCACTGCTGGAGTTGGAAAGCTGGAGTTTTTATTTGCTCTG 6720  
DB |||||||  
QY 6661 ACACCTATGATGTTTCACTGCACTGCTGGAGTTGGAAAGCTGGAGTTTTTATTTGCTCTG 6720  
DB |||||||  
QY 6721 GACCATTAAACACATATTAATGACCATGATTTTGTGGATATATATGACATAGTAGCT 6780  
DB |||||||  
QY 6721 GACCATTAAACACATATTAATGACCATGATTTTGTGGATATATATGACATAGTAGCT 6780  
DB |||||||

QY 6781 GAACTGAGAAGTGAAGAATGTGCATGTGTGCAAGATCTGGCACAGTATATCTTTTACAC 6840  
DB |||||||  
QY 6781 GAACTGAGAAGTGAAGAATGTGCATGTGTGCAAGATCTGGCACAGTATATCTTTTACAC 6840  
DB |||||||  
QY 6841 CAGTGCATTTCTGGATCTCTTATCAAAATAAGGGAAGTAAATCAGCCCATCTGTTTGTAAAC 6900  
DB |||||||  
QY 6841 CAGTGCATTTCTGGATCTCTTATCAAAATAAGGGAAGTAAATCAGCCCATCTGTTTGTAAAC 6900  
DB |||||||  
QY 6901 TATTTCAGCACTTCAGAAAGATGGACTCTTTTGGACGCCATGGAGGTGATGTTGAGCTTGAA 6960  
DB |||||||  
QY 6901 TATTTCAGCACTTCAGAAAGATGGACTCTTTTGGACGCCATGGAGGTGATGTTGAGCTTGAA 6960  
DB |||||||  
QY 6961 TGGGAAGAAAACCACTATGTAAATATTTCAGACCAAAAGGATACAAATTCGGAAGATTTTAA 7020  
DB |||||||  
QY 6961 TGGGAAGAAAACCACTATGTAAATATTTCAGACCAAAAGGATACAAATTCGGAAGATTTTAA 7020  
DB |||||||  
QY 7021 ATCCAGGGGCCAAAAGTTACCCCTCATTTCCGAAATGAAATGTGCACCTTAAAGAA 7080  
DB |||||||  
QY 7021 ATCCAGGGGCCAAAAGTTACCCCTCATTTCCGAAATGAAATGTGCACCTTAAAGAA 7080  
DB |||||||  
QY 7081 ATATCTATGCTTCTCTCACTGTGCCCTTT 7108  
DB |||||||  
QY 7081 ATATCTATGCTTCTCTCACTGTGCCCTTT 7108  
DB |||||||

RESULT 3  
US-10-466-759-10  
; Sequence 10, Application US/10466759  
; Publication No. US20040081983A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE CORPORATION  
; APPLICANT: LEE, Ernestine A.  
; APPLICANT: CHAWLA, Narinder K.  
; APPLICANT: BAUGHN, Mariah R.  
; APPLICANT: ISON, Craig H.  
; APPLICANT: GURURAJAN, Rajagopal  
; APPLICANT: ARVIZU, Chandra S.  
; APPLICANT: YAO, Monique G.  
; APPLICANT: JACKSON, Jennifer L.  
; APPLICANT: TANG, Y. Tom  
; APPLICANT: YUE, Henry  
; APPLICANT: TRAN, Bao  
; APPLICANT: DING, Li  
; APPLICANT: LU, Dying Aina M.  
; APPLICANT: LAL, Preeti G.  
; APPLICANT: WARREN, Bridget A.  
; TITLE OF INVENTION: KINASES AND PHOSPHATASES  
; FILE REFERENCE: PI-0344 USN  
; CURRENT APPLICATION NUMBER: US/10/466,759  
; CURRENT FILING DATE: 2003-07-17  
; PRIOR APPLICATION NUMBER: PCT/US02/01369  
; PRIOR FILING DATE: 2002-01-16  
; PRIOR APPLICATION NUMBER: US 60/263,083  
; PRIOR FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: US 60/271,205  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: US 60/271,117  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: US 60/276,859  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: US 60/278,504  
; PRIOR FILING DATE: 2001-03-23  
; PRIOR APPLICATION NUMBER: US 60/278,522  
; PRIOR FILING DATE: 2001-03-23  
; PRIOR APPLICATION NUMBER: US 60/280,510  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: US 60/280,266  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PERL Program  
; SEQ ID NO 10  
; LENGTH: 7225  
; TYPE: DNA





2128 ATTGATGTTTACCCGACATGAATAAAGGTTGAAGTGTGTCACCCGAAAAAGCCCAATGGG 2187  
2023 ATTGATGTTTACCCGACATGAATAAAGGTTGAAGTGTGTCACCCGAAAAAGCCCAATGGG 2082  
2188 ATCAATTAATGCTTTATGAAGTGTATATAAATAATATAGATACATTTATATATGAAGAACA 2247  
2083 ATCAATTAATGCTTTATGAAGTGTATATAAATAATATAGATACATTTATATATGAAGAACA 2142  
2248 TCAACAACAGACATAATATTAAGGAACTTAAGACCTCACACCTCTATACACATTTCTGTA 2307  
2143 TCAACAACAGACATAATATTAAGGAACTTAAGACCTCACACCTCTATACACATTTCTGTA 2202  
2308 AGGCTTTACACAGATTTGGTCAATGGCAATCAGGTATCTTTTACTCTCTGTAAGGACT 2367  
2203 AGGCTTTACACAGATTTGGTCAATGGCAATCAGGTATCTTTTACTCTCTGTAAGGACT 2262  
2368 TCGGAGCTGTGCTGATAGTGACCCAGAAATATACATTAACAAATATTTCTTCTGGA 2427  
2263 TCGGAGACTGTGCTGATAGTGACCCAGAAATATACATTAACAAATATTTCTTCTGGA 2322  
2428 GAGATTGAGCTATCATTTCTTCCCAAGTGTGTCACCAATGGCAATCAATAAATAATATACA 2487  
2323 GAGATTGAGCTATCATTTCTTCCCAAGTGTGTCACCAATGGCAATCAATAAATAATATACA 2382  
2488 ATTTATCTCAAGAGAAGTAATGGAATGAGGAAAGAACTATATAACAACCTCTTTAAAC 2547  
2383 ATTTATCTCAAGAGAAGTAATGGAATGAGGAAAGAACTATATAACAACCTCTTTAAAC 2442  
2548 CAACAATTAAGTACTGAAGAAATATACCAATATATCATTTGAGGTGTCTGCTAGTACA 2607  
2443 CAACAATTAAGGACTGAAGAAATATACCAATATATCATTTGAGGTGTCTGCTAGTACA 2502  
2608 CTGAAGAGTGAAGGAGTCTCGAGTGTCTCCATAAGTATACTGACGGAGGAAGTGTCTCT 2667  
2503 CTGAAGAGTGAAGGAGTCTCGAGTGTCTCCATAAGTATACTGACGGAGGAAGTGTCTCT 2562  
2668 GATTCTCCCTCTCAAGACTTCTCTGTAACACAGTGTCTGCTGTGTCTACCGTGAAGTTGTCA 2727  
2563 GATTCTCCCTCTCAAGACTTCTCTGTAACACAGTGTCTGCTGTGTCTACCGTGAAGTTGTCA 2622  
2728 TGGCAACACCCCTGGAGCAAAATGGAATATCTTTTATATACAGTTTATGTCT---GG 2784  
2623 TGGCAACACCCCTGGAGCAAAATGGAATATCTTTTATATACAGTTTATGTCTGTGGAGG 2682  
2785 AATAGATCATCATTAATAAACTATTAATGTCTGAAACATCATTTGGAGTTATCAGATTTG 2844  
2683 AATAGATCATCATTAATAAACTATTAATGTCTGAAACATCATTTGGAGTTATCAGATTTG 2742  
2845 GATTATAATGTTGAATACAGTGTCTTATGTAACAGCTAGCACACAGATTTGGTGTATGGAAA 2904  
2743 GATTATAATGTTGAATACAGTGTCTTATGTAACAGCTAGCACACAGATTTGGTGTATGGAAA 2802  
2905 ACAGGAAGCAATATCATTAATAGCTTTTCAAAACAGAGGAGAGCCAAAGCATCTCTCCAAA 2964  
2803 ACAAGGAAGCAATATCATTAATAGCTTTTCAAAACAGAGAGG---ACCAAGCGATCTCTCCAAA 2859  
2965 GATGTTTATATGCAACCTCAGTCTCTCATTAATAATTTCTTTCTGGACACCTCTCTCA 3024  
2860 GATGTTTATATGCAACCTCAGTCTCTCATTAATAATTTCTTTCTGGACACCTCTCTCA 2919  
3025 AAACCTAATGGGATTAATAAATTAATCTCTGTTTATTAAGAAATATCTTTCAGGTACTTTT 3084  
2920 AAACCTAATGGGATTAATAAATTAATCTCTGTTTATTAAGAAATATCTTTCAGGTACTTTT 2979  
3085 ATGCAGAAATTTTACCTCCATGAACATAACATGACTTTTGAACAATATGATCTGTATCCACA 3144  
2980 ATGCAGAAATTTTACCTCCATGAAGTAACCAATGACTTTTGAACAATATGATCTGTATCCACA 3039  
3145 ATTTATAGATAAATCTGCAATATTTACGTACTACTATATTTTGGTTTAAACAGCAAGTACTTCA 3204  
3040 ATTTATAGATAAATCTGCAATATTTACGTACTACTATATTTTGGTTTAAACAGCAAGTACTTCA 3099  
3205 GTTGGAAATGGGAAATAAAAGCAGTGTACATCATTTGAAGTATACACAGATCAAGACATACCT 3264

3100 GTTGGAAATGGGAAATAAAAGCAGTGTACATCATTTGAAGTATACACAGATCAAGACGTACCT 3159  
3265 GAAGGTTTGTGTTGGAAACCTGACTTACGNAATCCATTTGCTCACTGCRATTAATCTTAAGC 3324  
3160 GAAGGTTTGTGTTGGAAACCTGACTTACGNAATCCATTTGCTCACTGCRATTAATCTTAAGC 3219  
3325 TGGGTCCCAACCGGCTCAACCAACCGTCTAGTCTTCTACTATGTTTCTACTGATCTTACAG 3384  
3220 TGGGTCCCAACCGGCTCAACCAACCGTCTAGTCTTCTACTATGTTTCTACTGATCTTACAG 3279  
3385 CAGACTCTCTCGCATGTGAGACCACTCTTGTGTACATATGAGAGAAGCATATATTTTGTAT 3444  
3280 CAGACTCTCTCGCATGTGAGACCACTCTTGTGTACATATGAGAGAAGCATATATTTTGTAT 3339  
3445 AATCTGGAATAATACACTGATTTATATATTAATAAATTTACTCCATCAACAGAAAAGGATTC 3504  
3340 AATCTGGAATAATACACTGATTTATATATTAATAAATTTACTCCATCAACAGAAAAGGATTC 3399  
3505 TCTGATACCTTACTTGCCTGAGCTATACATCAAGACTGGAAGAAGATGTGCCAGAACTTCA 3564  
3400 TCTGATACCTTACTTGCCTGAGCTATACATCAAGACTGGAAGAAGATGTGCCAGAACTTCA 3459  
3565 CCAATTAATCAACACTTTTAAAAACCTTCTCTACTCTCACTGTTCTCTTATCATGGATCCC 3624  
3460 CCAATTAATCAACACTTTTAAAAACCTTCTCTACTCTCACTGTTCTCTTATCATGGATCCC 3519  
3625 CCAGTAAGCCAAATGGTGCAATATAGTTATGATTTAACTTTCAAGGACCAATGAA 3684  
3520 CCAGTAAGCCAAATGGTGCAATATAGTTATGATTTAACTTTCAAGGACCAATGAA 3579  
3685 AATTATTTCTTTCTCATTTACTTCTGATTAATACATAATATTGGAAGAGCTTTTCCACATTACA 3744  
3580 AATTATTTCTTTCTCATTTACTTCTGATTAATACATAATATTGGAAGAGCTTTTCCACATTACA 3639  
3745 TTATATAGCTTTTCTGCTCCCAAGAACTAGAAAAGGACTTTGGTCTTCCAGTATTTCT 3804  
3640 TTATATAGCTTTTCTGCTCCCAAGAACTAGAAAAGGACTTTGGTCTTCCAGTATTTCT 3699  
3805 TTCTTTTACACAGATGAGTCACTGCGGTAGCACCTCCACAAAATTTGACTTTTAACTCAAC 3864  
3700 TTCTTTTACACAGATGAGTCACTGCGGTAGCACCTCCACAAAATTTGACTTTTAACTCAAC 3759  
3865 TGCTACTCAGACTTTGATGCTGAAATGGAGCCCAAGTCTCTTCCAGTGTGTTATGTT 3924  
3760 TGCTACTCAGACTTTGATGCTGAAATGGAGCCCAAGTCTCTTCCAGTGTGTTATGTT 3819  
3925 AAAGTATATAGTTTAAAAATTCATGAAACATGAAACTGACACTATATATTAAGAATATA 3984  
3820 AAAGTATATAGTTTAAAAATTCATGAAACATGAAACTGACACTATATATTAAGAATATA 3879  
3985 TCAGATTTAAAACTGAAGCCAAACTTGTGGACGTGGAACCACTGACGACCTACTCTATC 4044  
3880 TCAGATTTAAAACTGAAGCCAAACTTGTGGACGTGGAACCACTGACGACCTACTCTATC 3939  
4045 CGTGATCTGCGTTCCACAAAGTGTGAAATGGCAATCAATTTAGTAATGTAGTAAATTC 4104  
3940 CGTGATCTGCGTTCCACAAAGTGTGGAATGGCAATCAATTTAGTAATGTAGTAAATTC 3999  
4105 ACAACCCCAAGATCAGTTTCCAGATGCTGTGCAAGATATGCAGTGCATGGCAACTAGCTGG 4164  
4000 ACAACCCCAAGATCAGTTTCCAGATGCTGTGCAAGATATGCAGTGCATGGCAACTAGCTGG 4059  
4165 CAGTCACTGTTTGTGAATGGATCCACCCAAAGGCAATGGAATTAATAACGCAATAT 4224  
4060 CAGTCACTGTTTGTGAATGGATCCACCCAAAGGCAATGGAATTAATAACGCAATAT 4119  
4225 ATGGTAAACAGTTGAAAGGAAATTTCTACAAAAGTTTCTCCCAAGATCAATGTACACTTTC 4284  
4120 ATGGTAAACAGTTGAAAGGAAATTTCTACAAAAGTTTCTCCCAAGATCAATGTACACTTTC 4179  
4285 ATAAAGCTTCTTGGCAATACCTCATATGTTCTTTAAAGTAAGAGCTTCAACCTCAGCTGGT 4344







|    |      |   |      |
|----|------|---|------|
| Qy | 2911 | AGCAATATCATTTAGCTTTTCAAACACACAGAGGAGCACCAAGCGATCTCTCCAAAGATGTT    | 2970 |
| Db | 2812 | AGCAATATCATTTAGCTTTTCAAACACACAGAGGG--ACCAAGCGATCTCTCCAAAGATGTT    | 2868 |
| Qy | 2971 | TATTATGCAAAACCTTCAGTCTTTCATCAATTAATTTCTTTTCTGGGACACCTCTTCAAAACCT  | 3030 |
| Db | 2869 | TATTATGCAAAACCTTCAGTCTTTCATCAATTAATTTCTTTTCTGGGACACCTCTTCAAAACCT  | 2928 |
| Qy | 3031 | AATGGGATTAACAATTAATCTCTGTTTATTAACAGAAATACTTTCAGTACTTTTATGTCAG     | 3090 |
| Db | 2929 | AATGGGATTAACAATTAATCTCTGTTTATTAACAGAAATACTTTCAGTACTTTTATGTCAG     | 2988 |
| Qy | 3091 | AATTTTACACTCCATGAACCTAAACCAATGACTTTTGACAAATAGACTGTATCCAAATTAATA   | 3150 |
| Db | 2989 | AAATTTACACTCCATGAAGTAAACCAATGACTTTTGACAAATAGACTGTATCCAAATTAATA    | 3048 |
| Qy | 3151 | GATAAACTGCAATATTTTCAGCTACTATACATATTTTGGTTAAACAGCAAGTACTTCAGTTGGGA | 3210 |
| Db | 3049 | GATAAACTGCAATATTTTCAGCTACTATACATATTTTGGTTAAACAGCAAGTACTTCAGTTGGGA | 3108 |
| Qy | 3211 | AATGGGAATAAAGCAGTGACATCAATTTGAAGTATACACAGATCAAGACATACCTGAGGG      | 3270 |
| Db | 3109 | AATGGGAATAAAGCAGTGACATCAATTTGAAGTATACACAGATCAAGACATACCTGAGGG      | 3168 |
| Qy | 3271 | TTTGTGTGAAAACCTGACTTACGAATCCAAATTCGTCAACTGCAATAAATGTAAAGCTGGGTC   | 3330 |
| Db | 3169 | TTTGTGTGAAAACCTGACTTACGAATCCAAATTCGTCAACTGCAATAAATGTAAAGCTGGGTC   | 3228 |
| Qy | 3331 | CCACCGGCTCAACCAAAACGGTCTAGTCTTCTACTATGTTTTCAGTGTCTTTACAGCAGACT    | 3390 |
| Db | 3229 | CCACCGGCTCAACCAAAACGGTCTAGTCTTCTACTATGTTTTCAGTGTCTTTACAGCAGACT    | 3288 |
| Qy | 3391 | CTTCGCCATGTGAGACCACTCTTGTTTACATATGAGAGAAGCATATATTTTGTATATCTG      | 3450 |
| Db | 3289 | CCTCGCCATGTGAGACCACTCTTGTTTACATATGAGAGAAGCATATATTTTGTATATCTG      | 3348 |
| Qy | 3451 | GAATAATACACTGATTAATATATTAATAAATTTACTTCCATCAACAGAAAAGGATTTCTCTGAT  | 3510 |
| Db | 3349 | GAATAATACACTGATTAATATATTAATAAATTTACTTCCATCAACAGAAAAGGATTTCTCTGAT  | 3408 |
| Qy | 3511 | ACCTATATCTGCCAGCTATATACATCAAGACTGAAGAAGATGTCCAGAAAACCTTCAACCAATA  | 3570 |
| Db | 3409 | ACCTATATCTGCCAGCTATATACATCAAGACTGAAGAAGATGTCCAGAAAACCTTCAACCAATA  | 3468 |
| Qy | 3571 | ATCAACACTTTTAAAAAACCTTTCCTCTACCTCAGTTCCTTTATCATGGGATCCCCAGTA      | 3630 |
| Db | 3469 | ATCAACACTTTTAAAAAACCTTTCCTCTACCTCAGTTCCTTTATCATGGGATCCCCAGTA      | 3528 |
| Qy | 3631 | AAGCCAAATGGTGCAATAATAAGTTATGATTTAACTTTTACAAGGACCAAAATGAATAATAT    | 3690 |
| Db | 3529 | AAGCCAAATGGTGCAATAATAAGTTATGATTTAACTTTTACAAGGACCAAAATGAATAATAT    | 3588 |
| Qy | 3691 | TCCTTTCATTAATCTCTGATAATTAACAATAATTTGGAAGAGCTTTTCAACCAATTTACATTAAT | 3750 |
| Db | 3589 | TCCTTTCATTAATCTCTGATAATTAACAATAATTTGGAAGAGCTTTTCAACCAATTTACATTAAT | 3648 |
| Qy | 3751 | AGCTTTTTCCTGTCGCGCAAGAACTAGAAAAAGGACTTGGTCTCTTCAGTATTCCTTTCTTTT   | 3810 |
| Db | 3649 | AGCTTTTTCCTGTCGCGCAAGAACTAGAAAAAGGACTTGGTCTCTTCAGTATTCCTTTCTTTT   | 3708 |
| Qy | 3811 | TACACAGATGATCAGTGGCGTTAGCACCTTCCACAAAATTTGACTTTTAATCAACTGTACT     | 3870 |
| Db | 3709 | TACACAGATGATCAGTGGCGTTAGCACCTTCCACAAAATTTGACTTTTAATCAACTGTACT     | 3768 |
| Qy | 3871 | TCAGACTTTGTATGGCTGAAATGAGGCCCAAGTCTCTTCAGGTGGTATTTGTTAAAGTA       | 3930 |
| Db | 3769 | TCAGACTTTGTATGGCTGAAATGAGGCCCAAGTCTCTTCAGGTGGTATTTGTTAAAGTA       | 3828 |
| Qy | 3931 | TATAGTTTTTAAAAATTCATGAACATGAACCTGACACTATATATTAATAAGATATATCAGGA    | 3990 |
| Db | 3829 | TATAGTTTTTAAAAATTCATGAACATGAACCTGACACTATATATTAATAAGATATATCAGGA    | 3888 |

[illegible]









Db 1498 AACTTCTCCTGCAAGGAATAGAGCTGAAGACGACCTTCCAGCTTGTAACTACAAGAAAT 1557  
Qy 1660 CAGTATATTAATGACATTTGAGCTGAACAGCTGTCTTATGTATTCAGGAGACTTGTACCT 1719  
Db 1558 CAGTATATTAATGACATTTGAGCTGAACAGCTGTCTTATGTATTCAGGAGACTTGTACCT 1617  
Qy 1720 TTCACTGAGCACAATGATAGTGTATCTGCTTTCACCATCATGGAGAGGACCAACACA 1779  
Db 1618 TTCACTGAGCACAATGATAGTGTATCTGCTTTCACCATCATGGAGAGGACCAACACA 1677  
Qy 1780 GTTCTCAGTGTAGGACACGCTCAGCAAGTGCCTCAAGCTCCATTAATAATTAATAACTATAA 1839  
Db 1678 GTTCTCAGTGTAGGACACGCTCAGCAAGTGCCTCAAGCTCCATTAATAATTAATAACTATAA 1737  
Qy 1840 AATATTAGTTCCTTCATCTAATTTTGTATATTTGGATCTCTCCAGAAATATCCCAATGGAAA 1899  
Db 1738 AATATTAGTTCCTTCATCTAATTTTGTATATTTGGATCTCTCCAGAAATATCCCAATGGAAA 1797  
Qy 1900 ATAACTCATTATACGATTTATGCAATGGAAATTTGGATACAAACAGAGGATTCAGATACT 1959  
Db 1798 ATAACTCATTATACGATTTATGCAATGGAAATTTGGATACAAACAGAGCAATTCAGATACT 1857  
Qy 1960 ACCATAGATAACAGCTTCTCTATAAC-----AGGGTTAAAGAAATACACAAATAACAAA 2013  
Db 1858 ACCATAGATAACAGCTTCTCTATAACAGGTATAGGGTTAAAGAAATACACAAATAACAAA 1917  
Qy 2014 ATGAGAGTGGCAGCTCAACCCAGATGGAGAAAGTTCTTTGTCTGAAGAAATGACATC 2073  
Db 1918 ATGAGAGTGGCAGCTCAACCCAGGTGGAGAAAGTTCTTTGTCTGAAGAAATGACATC 1977  
Qy 2074 TTTGTGAGAACTTCAGAGATGAACCCGGAATCATCACTCAAGATGTGGAAGTAAATGAT 2133  
Db 1978 TTTGTGAGAACTTCAGAGATGAACCCGGAATCATCACTCAAGATGTGGAAGTAAATGAT 2037  
Qy 2134 GTTACCGCAGATCAAAATAGTTGAAGTGTGACACCCGGAAGCCCAATGGATCATTT 2193  
Db 2038 GTTACCGCAGATCAAAATAGTTGAAGTGTGACACCCGGAAGCCCAATGGATCATTT 2097  
Qy 2194 ATTGCTTATGAAGTGTATATAAAATATAGATACATTTATATGAAGAAACATCAACA 2253  
Db 2098 ATTGCTTATGAAGTGTATATAAAATATAGATACATTTATATGAAGAAACATCAACA 2157  
Qy 2254 ACAGACATATATTAAGGAACCTTAAGACCTCAACCCCTCTATAAACATTTCTGPAAGGTCT 2313  
Db 2158 ACAGACATATATTAAGGAACCTTAAGACCTCAACCCCTCTATAAACATTTCTGPAAGGTCT 2217  
Qy 2314 TACACAGATTTGGTCAATGCGCAATCAGGTATCTTCTTACTCTCTGTAAGGACTTCGGAG 2373  
Db 2218 TACACAGATTTGGTCAATGCGCAATCAGGTATCTTCTTACTCTCTGTAAGGACTTCGGAG 2277  
Qy 2374 ACTGTGCTGATAGTGCACAGAAATATACCTTACAAATAATTTCTTCTGAGAGATT 2433  
Db 2278 ACTGTGCTGATAGTGCACAGAAATATACCTTACAAATAATTTCTTCTGAGAGATT 2337  
Qy 2434 GAGCTATCATTTCTTCCCCAGTAGTCCCAATGGAATCATAAAAAATATACAATTTAT 2493  
Db 2338 GAGCTATCATTTCTTCCCCAGTAGTCCCAATGGAATCATAAAAAATATACAATTTAT 2397  
Qy 2494 CTCAGAGAGATTAATGGAATGAGGAAGAACTATATAAATACACCTCTTTAAACCCCAAAAC 2553  
Db 2398 CTCAGAGAGATTAATGGAATGAGGAAGAACTATATAAATACACCTCTTTAAACCCCAAAAC 2457  
Qy 2554 ATTTAAAGTACTGAAGAAATATACCAATATATCATTTAGGTGTCTGTAGTACACTGAAA 2613  
Db 2458 ATT-----CTGAAGAAATATACCAATATATCATTTAGGTGTCTGTAGTACACTCAAA 2511  
Qy 2614 GGTGAAGGAGTTCGGAGTGTCTCCATAAGTATATCTGACGGAGGAAGATGCTCTGTATCT 2673  
Db 2512 GGTGAAGGAGTTCGGAGTGTCTCCATAAGTATATCTGACGGAGGAAGATGCTCTGTATCT 2571  
Qy 2674 CCCCCCTCAAGACTTCTCTGTAAACAGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2733  
Db 2572 CCCCCCTCAAGACTTCTCTGTAAACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2631

Qy 2734 CCACCCCTGGAGCCAAATGGAATTAATCCTTTTATTAACAGTTTATGTCT---GGAATAGA 2790  
Db 2632 CCACCCCTGGAGCCAAATGGAATTAATCCTTTTATTAACAGTTTATGTCTGTGGAGAAATAGA 2691  
Qy 2791 TCATCAATTAATAAATTAATTAATGTCTCACTGAAAACATCATTTGGAGTTATCAGATTTGGATAT 2850  
Db 2692 TCATCAATTAATAAATTAATTAATGTCTCACTGAAAACATCATTTGGAGTTATCAGATTTGGATAT 2751  
Qy 2851 AATGTTGAATACAGTGTCTTATGTAACAGCTAGCACAGAAATTTGGTGTATGGAAAAACAGGA 2910  
Db 2752 AATGTTGAATACAGTGTCTTATGTAACAGCTAGCACAGAAATTTGGTGTATGGAAAAACAGGA 2811  
Qy 2911 AGCAATATCATTTAGCTTTCAAAACACAGAGGAGGACCAAGCGATCCTCCCAAGATGTT 2970  
Db 2812 AGCAATATCATTTAGCTTTCAAAACACAGAGG---ACCAAGCGATCCTCCCAAGATGTT 2868  
Qy 2971 TATTAATGCAAAACCTCAGTTCTTCATCAATAAATTTCTTTCTGACACCTCTCTCAAAACCT 3030  
Db 2869 TATTAATGCAAAACCTCAGTTCTTCATCAATAAATTTCTTTCTGACACCTCTCTCAAAACCT 2928  
Qy 3031 AATGGGATTAACAATATTAATCTCTGTTTATTAACAGAAATACCTTCAGGTACTTTTATGTCAG 3090  
Db 2929 AATGGGATTAACAATATTAATCTCTGTTTATTAACAGAAATACCTTCAGGTACTTTTATGTCAG 2988  
Qy 3091 AATTTTACCTCCATGAACCTAACCAATGACTTTTGACAAATATGATGTATCCCAATATATA 3150  
Db 2989 AATTTTACCTCCATGAAGTAACCAATGACTTTTGACAAATATGATGTATCCCAATATATA 3048  
Qy 3151 GATTAACGTGCAATATTAATGACTACTATATATTTGGTGTAAACAGCAAGTACTTCAGTTGGA 3210  
Db 3049 GATTAACGTGCAATATTAATGACTACTATATATTTGGTGTAAACAGCAAGTACTTCAGTTGGA 3108  
Qy 3211 AATGGGAATTAAGCAGTGCACATCATTTGAAGTATACACAGATCAACAGACATCTGGAAGGG 3270  
Db 3109 AATGGGAATTAAGCAGTGCACATCATTTGAAGTATACACAGATCAACAGACATCTGGAAGGG 3168  
Qy 3271 TTTGTTGGAAACCTGACTTTACGAATCCATTTCTGTCACCTGCAATTAATTAAGTCTGGTCT 3330  
Db 3169 TTTGTTGGAAACCTGACTTTACGAATCCATTTCTGTCACCTGCAATTAATTAAGTCTGGTCT 3328  
Qy 3331 CCACGGCTCAACCAAAACCGTCTAGTCTTCTACTATGTTTCTACTGATCTTTACAGCAGACT 3390  
Db 3229 CCACGGCTCAACCAAAACCGTCTAGTCTTCTACTATGTTTCTACTGATCTTTACAGCAGACT 3288  
Qy 3391 CCTCGCATGTGAGACCACTCTGTTACATATGAGAGAGCATATATTTTGTATATCTG 3450  
Db 3289 CCTCGCATGTGAGACCACTCTGTTACATATGAGAGAGCATATATTTTGTATATCTG 3348  
Qy 3451 GAAAAATACACTGATTATATATTAATAAATTTACTCCATCAACAGAAAGGATTTCTCTGAT 3510  
Db 3349 GAAAAATACACTGATTATATATTAATAAATTTACTCCATCAACAGAAAGGATTTCTCTGAT 3408  
Qy 3511 ACCTATATGCCCCAGCTATATACATCAAGACTGAAGAGATGTCCAGAAAACCTTCACCAATA 3570  
Db 3409 ACCTATATGCCCCAGCTATATACATCAAGACTGAAGAGATGTCCAGAAAACCTTCACCAATA 3468  
Qy 3571 ATCAACACTTTTAAAAACCTTTCTCTACCTCAGTCTCTTATCATGGATCCCCAGTA 3630  
Db 3469 ATCAACACTTTTAAAAACCTTTCTCTACCTCAGTCTCTTATCATGGATCCCCAGTA 3528  
Qy 3631 AAGCCAAATGTTGCAATAATAGTTATGATTTAACTTTTCAAGGACCAATGAAAAATTAAT 3690  
Db 3529 AAGCCAAATGTTGCAATAATAGTTATGATTTAACTTTTCAAGGACCAATGAAAAATTAAT 3588  
Qy 3691 TCTTTTCACTTCTGTGAATAATATATATTTGGAAGAGCTTTCCACCAATATATATAT 3750  
Db 3589 TCTTTTCACTTCTGTGAATAATATATATTTGGAAGAGCTTTCCACCAATATATATAT 3648  
Qy 3751 AGCTTTTGTGCGCGCAAGACTAGAAAGCACTTGGTCTTCCAGTATCTCTTCTTT 3810  
Db 3649 AGCTTTTGTGCGCGCAAGACTAGAAAGCACTTGGTCTTCCAGTATCTCTTCTTT 3708

3811 TACACAGATGAGTCAGTGCCTGTTAGCACTCCACAAAATTTGACATTTTAATCAACTGTACT 3870  
3709 TACACAGATGAGTCAGTGCCTGTTAGCACTCCACAAAATTTGACATTTTAATCAACTGTACT 3768  
3871 TCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGGTATTTGTTAAAGTA 3930  
3769 TCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGGTATTTGTTAAAGTA 3828  
3931 TATAGTTTTTAAATTCATGAACATGAATGCACTGATATATATATATAAGAAATATATACGGA 3990  
3829 TATAGTTTTTAAATTCATGAACATGAATGCACTGATATATATATAAGAAATATATACGGA 3888  
3991 TTTTAAACTGAAGCCAAACTTTGTGACCTGGAAACAGTCAGCACCTACTCTATCCGTGTA 4050  
3889 TTTTAAACTGAAGCCAAACTTTGTGACCTGGAAACAGTCAGCACCTACTCTATCCGTGTA 3948  
4051 TCTGCGTTCCAAAGTTGGAATGGCAATCAATTTAGTAAATTTAGTAAATTTCAACCC 4110  
3949 TCTGCGTTCCAAAGTTGGAATGGCAATCAATTTAGTAAATTTAGTAAATTTCAACCC 4008  
4111 CAAGATCAGTTCCAGATGCTGTCAGAAATATGCACTGATGCACTGGCACTAGCTGGCAGTCA 4170  
4009 CAAGATCAGTTCCAGATGCTGTCAGAAATATGCACTGATGCACTGGCACTAGCTGGCAGTCA 4068  
4171 GTTTTAGTGAATGGGATCCACCCAAAAGGCAATGGAATTAATTAACGAGTATATGTA 4230  
4069 GTTTTAGTGAATGGGATCCACCCAAAAGGCAATGGAATTAATTAACGAGTATATGTA 4128  
4231 ACAGTTGAAAGGAATTTCTACAAAGTTTCTCCCAAGATCAATGTACATCTTTCTATAAAG 4290  
4129 ACAGTTGAAAGGAATTTCTACAAAGTTTCTCCCAAGATCAATGTACATCTTTCTATAAAG 4188  
4291 CTTCTGCGCAATACCTCATATGCTTTTAAAGTAAGAGCTTCAACCTCAGCTGGTGAAGGT 4350  
4189 CTTCTGCGCAATACCTCATATGCTTTTAAAGTAAGAGCTTCAACCTCAGCTGGTGAAGGT 4248  
4351 GATGAAAGCACATGCCATGTCAGACACATACCTGAAACAGTTCCCAAGTGTCCCACAAT 4410  
4249 GATGAAAGCACATGCCATGTCAGACACATACCTGAAACAGTTCCCAAGTGTCCCACAAT 4308  
4411 ATTGCTTTTTCTGATGTTCACTGATGTCAGTGAACATGTCAGATGGAATAGACTGACACT 4470  
4309 ATTGCTTTTTCTGATGTTCACTGATGTCAGTGAACATGTCAGATGGAATAGACTGACACT 4368  
4471 ATCTCTGGCTACTTTTCAAAATTTACAAAATTTACCACTCAACTCTGCTCCAAAATGCAAA 4530  
4369 ATCTCTGGCTACTTTTCAAAATTTACAAAATTTACCACTCAACTCTGCTCCAAAATGCAAA 4428  
4531 GAATGGGAATCCGAAGAATGTGTTGAATATCAAAAATTTCAATACCTCTATGAAGCTCAC 4590  
4429 GAATGGGAATCCGAAGAATGTGTTGAATATCAAAAATTTCAATACCTCTATGAAGCTCAC 4488  
4591 TTAACCTGAAGACAGTATATGGATTAAGAAATTTAGATGGTATAGATTCOAAGTGGCT 4650  
4489 TTAACCTGAAGACAGTATATGGATTAAGAAATTTAGATGGTATAGATTCOAAGTGGCT 4548  
4651 GCCAGCACCAATGCTGGCTATGGCAATGCTTCAACTGATGTTCTACAAAATCTGCGCT 4710  
4549 GCCAGCACCAATGCTGGCTATGGCAATGCTTCAACTGATGTTCTACAAAATCTGCGCT 4608  
4711 GGCCCTCCAGATGGTCTCTGAAAATGTTCAATGATAGTACCAATCACCTTTTATGATC 4770  
4609 GGCCCTCCAGATGGTCTCTGAAAATGTTCAATGATAGTACCAATCACCTTTTATGATC 4668  
4771 AGCATAAGCTGGAGTGAACCTGCTGTCATTACTGGAACCAATGTTATCTGATGATGTC 4830  
4669 AGCATAAGCTGGAGTGAACCTGCTGTCATTACTGGAACCAATGTTATCTGATGATGTC 4728  
4831 AAATCGGTAGATAAGTAAATTTATATATATCTTCACTCACTCAATGCAATGAAGAAATATAA 4890  
4729 AAATCGGTAGATAAGTAAATTTATATATATCTTCACTCACTCAATGCAATGAAGAAATATAA 4788  
4891 ACCATAGAAATTTAAGAAATTTAGAAATTTATTTCAAGGTAATCTGTAGTATCTGATTT 4950

4789 ACCATAGAAATTTAAGAAATTTAGAAATTTTCAAGGTAATCTGTAGTATCTGATCACTGCAATTT 4848  
4951 ACTGGGACATTTAGTGTCTGCATATCTAGAAAGGAAGTCAAGTGTGAAATGATTTGTTACT 5010  
4849 ACTGGGACATTTAGTGTCTGCATATCTAGAAAGGAAGTCAAGTGTGAAATGATTTGTTACT 4908  
5011 ACTTTAGAAATCAGCCCCCAAGGACCCACCTAACCAACATGACATTTTCAGAAGATACAGAT 5070  
4909 ACTTTAGAAATCAGCCCCCAAGGACCCACCTAACCAACATGACATTTTCAGAAGATACAGAT 4968  
5071 GAAGTTACAAAATTTCAATTAACGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5130  
4969 GAAGTTACAAAATTTCAATTAACGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5028  
5131 GTATATCAAGCTCTGTTTACCGAGAAGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5190  
5029 GTATATCAAGCTCTGTTTACCGAGAAGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5088  
5191 AGTATTTATACAGAAACCAACACATCT 5250  
5089 AGTATTTATACAGAAACCAACACATCT 5148  
5251 CATACATCAATATCAGTGTCTTACCGAGTCAATAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5310  
5149 CATACATCAATATCAGTGTCTTACCGAGTCAATAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5208  
5311 ATGAGATAACCATGATATCAAGCTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5370  
5209 ATGAGATAACCATGATATCAAGCTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5268  
5371 TATGATGCCACAGGAAACTGCTGTGACTTCAACCAACAAATTAACAATCAGAAATGCAATA 5430  
5269 TATGATGCCACAGGAAACTGCTGTGACTTCAACCAACAAATTAACAATCAGAAATGCAATA 5328  
5431 TGTACTACAGTGTGATCATGGACCAATAAAAAATGTACAAGTCTTGTGTCAGAGAAC 5490  
5329 TGTACTACAGTGTGATCATGGACCAATAAAAAATGTACAAGTCTTGTGTCAGAGAAC 5388  
5491 GGAGCTCAGCATGATGGAATGTACAAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 5550  
5389 GGAGCTCAGCATGATGGAATGTACAAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 5448  
5551 CCATATTTTACAAATGAAGGCTTTCTTAAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5610  
5449 CCATATTTTACAAATGAAGGCTTTCTTAAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5508  
5611 AGTGCAATGAAGAAATCTCATCATAGTGTCTGATTAATGCAATGATGATGATGATGATGATGATGAT 5670  
5509 AGTGCAATGAAGAAATCTCATCATAGTGTCTGATTAATGCAATGATGATGATGATGATGATGATGAT 5568  
5671 GAAGCAAAATTTGCAATGGACCACTGAAACCAAAAAGCAATCTTATTTAAATTTAGA 5730  
5569 GAAGCAAAATTTGCAATGGACCACTGAAACCAAAAAGCAATCTTATTTAAATTTAGA 5628  
5731 GCTACAAATTTATCGGCAATTTACTGACTCTGATTAATTTCTGACCTCTGTTAAAGCTTTA 5790  
5629 GCTACAAATTTATCGGCAATTTACTGACTCTGATTAATTTCTGACCTCTGTTAAAGCTTTA 5688  
5791 GGGGAAGGACTTTTACAGAAAGAACCGTAGAGATCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5850  
5689 GGGGAAGGACTTTTACAGAAAGAACCGTAGAGATCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5748  
5851 TCATAATTTCTCTTGGAAACAGTATTTTGGCAATTTGGCAATTTGGCAATTTGGCAATTTGGCAATTT 5910  
5749 TCATAATTTCTCTTGGAAACAGTATTTTGGCAATTTGGCAATTTGGCAATTTGGCAATTTGGCAATTT 5808  
5911 GAAGGTGGCACAATCTCTCTCTCAGGATGCAAGAAATTTATTTGACACTTAAATTTGAAGCTGGAT 5970  
5809 GAAGGTGGCACAATCTCTCTCTCAGGATGCAAGAAATTTATTTGACACTTAAATTTGAAGCTGGAT 5868  
5971 CAGCTCATCAGAGTGGCAGACCTGGAACTGGAAGGACGAGAGATTAACCGCG 6020

|      |    |           |         |          |       |       |       |       |       |       |       |         |          |          |       |     |      |      |
|------|----|-----------|---------|----------|-------|-------|-------|-------|-------|-------|-------|---------|----------|----------|-------|-----|------|------|
| 5989 | DB | CAGCTCATC | CAGTGGC | AGACCTCG | GAAC  | CTGG  | GAAC  | CTGGA | ACTG  | GAAG  | CGAG  | GAGATTA | ACCGCGGT | TACTTAGT | 5992  |     |      |      |
| 6021 | QY | -----     | -----   | -----    | ----- | ----- | ----- | ----- | ----- | ----- | ----- | -----   | -----    | -----    | 6063  |     |      |      |
| 5929 | DB | TATAGAAA  | ATCCAT  | CAAGCC   | CAAT  | TAAG  | CAAG  | AAAT  | TCC   | TTC   | TG    | CAAC    | ATG      | TG       | AGAGC | TT  | 5988 |      |
| 6064 | QY | TGCA      | CAAC    | AC       | CAAC  | CTTA  | AG    | TTT   | TC    | AG    | AG    | AA      | TTT      | TC       | CG    | AA  | TT   | 6123 |
| 5989 | DB | TGCA      | CAAC    | CAAC     | CAAC  | CTTA  | AG    | TTT   | TC    | AG    | AG    | AA      | TTT      | TC       | CG    | AA  | TT   | 6048 |
| 6124 | QY | GAT       | CTT     | TCT      | CTT   | CAAA  | CT    | GA    | TG    | CT    | GC    | CTT     | TG       | GA       | AT    | AG  | CA   | 6183 |
| 6049 | DB | GAT       | CTT     | TCT      | CTT   | CAAA  | CT    | GA    | TG    | CT    | GC    | CTT     | TG       | GA       | AT    | AG  | CA   | 6108 |
| 6184 | QY | ATA       | AAA     | CC       | ATA   | TA    | TA    | TA    | TA    | TA    | TA    | TA      | TA       | TA       | TA    | TA  | TA   | 6243 |
| 6109 | DB | ATA       | AA      | CCAT     | ---   | ATA   | TA    | TA    | TA    | TA    | TA    | TA      | TA       | TA       | TA    | TA  | TA   | 6165 |
| 6244 | QY | TCG       | GA      | TTA      | TAT   | TAAT  | AT    | GC    | CA    | CT    | AT    | TAT     | TTC      | TG       | TTA   | TTT | AT   | 6303 |
| 6166 | DB | TCG       | GA      | TTA      | TAT   | TAAT  | AT    | GC    | CA    | CT    | AT    | TAT     | TTC      | TG       | TTA   | TTT | AT   | 6225 |
| 6304 | QY | ACT       | CA      | AG       | GT    | CCA   | CT    | TA    | CC    | AG    | GA    | CA      | AG       | T        | TG    | GA  | AT   | 6363 |
| 6226 | DB | ACT       | CA      | AG       | GT    | CCA   | CT    | TA    | CC    | AG    | GA    | CA      | AG       | T        | TG    | GA  | AT   | 6285 |
| 6364 | QY | GC        | AAA     | CA       | CT    | TAG   | TAA   | CT    | TA    | CA    | CA    | CT      | GT       | T        | TG    | AAA | AG   | 6423 |
| 6286 | DB | GC        | AAA     | CA       | CT    | TAG   | TAA   | CT    | TA    | CA    | CA    | CT      | GT       | T        | TG    | AAA | AG   | 6345 |
| 6424 | QY | TAT       | TG      | GC       | CA    | GAG   | CA    | CA    | CA    | GC    | CA    | GT      | T        | ACT      | TG    | CT  | TT   | 6483 |
| 6346 | DB | TAT       | TG      | GC       | CA    | GAG   | CA    | CA    | CA    | GC    | CA    | GT      | T        | ACT      | TG    | CT  | TT   | 6405 |
| 6484 | QY | ATG       | GA      | GA       | GT    | T     | CA    | AA    | T     | AG    | A     | TTG     | GA       | CT       | AT    | CA  | GG   | 6543 |
| 6406 | DB | ATG       | GA      | GA       | GT    | T     | CA    | AA    | T     | AG    | A     | TTG     | GA       | CT       | AT    | CA  | GG   | 6465 |
| 6544 | QY | TGC       | AT      | GA       | CT    | GT    | TC    | GA    | CA    | GT    | GT    | T       | ACT      | TG       | CT    | GC  | CT   | 6603 |
| 6466 | DB | TGC       | AT      | GA       | CT    | GT    | TC    | GA    | CA    | GT    | GT    | T       | ACT      | TG       | CT    | GC  | CT   | 6525 |
| 6504 | QY | AGC       | CC      | CT       | CT    | TA    | T     | CA    | CT    | TT    | TG    | GA      | AG       | T        | TG    | GA  | AG   | 6663 |
| 6526 | DB | AGC       | CC      | CT       | CT    | TA    | T     | CA    | CT    | TT    | TG    | GA      | AG       | T        | TG    | GA  | AG   | 6585 |
| 6664 | QY | CTT       | AT      | GA       | TT    | GT    | CT    | CA    | CT    | GA    | GT    | CT      | GA       | GT       | CT    | GA  | GT   | 6723 |
| 6586 | DB | CTT       | AT      | GA       | TT    | GT    | CT    | CA    | CT    | GA    | GT    | CT      | GA       | GT       | CT    | GA  | GT   | 6645 |
| 6724 | QY | CAT       | TT      | TA       | CA    | CA    | CA    | TAT   | ATA   | AT    | GA    | CC      | AT       | GA       | TT    | TG  | TG   | 6783 |
| 6546 | DB | CAT       | TT      | TA       | CA    | CA    | CA    | TAT   | ATA   | AT    | GA    | CC      | AT       | GA       | TT    | TG  | TG   | 6705 |
| 6784 | QY | CTG       | AG      | AG       | T     | G     | AA    | GA    | AT    | GT    | GC    | AG      | AT       | CT       | GC    | CA  | CA   | 6843 |
| 6706 | DB | CTG       | AG      | AG       | T     | G     | AA    | GA    | AT    | GT    | GC    | AG      | AT       | CT       | GC    | CA  | CA   | 6765 |
| 6844 | QY | TGC       | AT      | TT       | CT    | GA    | CT    | CT    | T     | AT    | CA    | AT      | TA       | CG       | CC    | CA  | CT   | 6903 |
| 6766 | DB | TGC       | AT      | TT       | CT    | GA    | CT    | CT    | T     | AT    | CA    | AT      | TA       | CG       | CC    | CA  | CT   | 6825 |
| 6904 | QY | TC        | AG      | CA       | CT    | TC    | AG    | AG    | AT    | TG    | GA    | CT      | CT       | TT       | TG    | GA  | CT   | 6963 |
| 6826 | DB | TC        | AG      | CA       | CT    | TC    | AG    | AG    | AT    | TG    | GA    | CT      | CT       | TT       |       |     |      |      |

RESULT 6  
US-10-08

```

1 Sequence 5, Application US/10087684
2 Publication No. US20040029116A1
3 GENERAL INFORMATION:
4 APPLICANT: Edinger, Shlomit R.
5 APPLICANT: MacDougall, John R.
6 APPLICANT: Millet, Isabelle
7 APPLICANT: Ellerman, Karen
8 APPLICANT: Stone, David J.
9 APPLICANT: Grosse, William M.
10 APPLICANT: Lepley, Denise M.
11 APPLICANT: Rieger, Daniel K.
12 APPLICANT: Burgess, Catherine E.
13 APPLICANT: Casman, Stacie, J.
14 APPLICANT: Spytek, Kimberly A.
15 APPLICANT: Boldog, Ferenc L.
16 APPLICANT: Li, Li
17 APPLICANT: Padigaru, Muralidhara
18 APPLICANT: Mishra, Vishnu
19 APPLICANT: Shenoy, Suresh G.
20 APPLICANT: Rastelli, Luca
21 APPLICANT: Tchernev, Velizar T.
22 APPLICANT: Vernet, Corine A.M.
23 APPLICANT: Zerhusen, Bryan D.
24 APPLICANT: Malyankar, Uriel M.
25 APPLICANT: Guo, Xiaojia
26 APPLICANT: Miller, Charles E.
27 APPLICANT: Gangolli, Esha A.
28 TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME
29 FILE REFERENCE: 21402-214 CIP
30 CURRENT APPLICATION NUMBER: US/10/087,684
31 CURRENT FILING DATE: 2003-03-10
32 PRIOR APPLICATION NUMBER: 60/253,834
33 PRIOR FILING DATE: 2000-11-29
34 PRIOR APPLICATION NUMBER: 60/250,926
35 PRIOR FILING DATE: 2000-11-30
36 PRIOR APPLICATION NUMBER: 60/264,180
37 PRIOR FILING DATE: 2001-01-25
38 PRIOR APPLICATION NUMBER: 60/274,194
39 PRIOR FILING DATE: 2001-03-08
40 PRIOR APPLICATION NUMBER: 60/313,656
41 PRIOR FILING DATE: 2001-08-20
42 PRIOR APPLICATION NUMBER: 60/327,456
43 PRIOR FILING DATE: 2001-10-05
44 NUMBER OF SEQ ID NOS: 220
45 SOFTWARE: CuraseqList version 0.1
46 SEQ ID NO 5
47 LENGTH: 6994
48 TYPE: DNA
49 ORGANISM: Homo sapiens
50 FEATURE:
51 NAME/KEY: CDS
52 LOCATION: (31)..(6873)
53 US-10-087-684-5

```

| Query Match           | 90.8%;          | Score 6453;  | DB 17;     | Length 6994; |
|-----------------------|-----------------|--|------------|--------------|
| Best Local Similarity | 96.3%;          | Pred. No. 0;   |            |              |
| Matches 6770;         | Conservative 0; | Mismatches 215;  | Indels 48; | Gaps 14      |
| Qy                    | 76              | TGATTCTACTGGCTGAAATAATGAATAAAGATGGATTCTTATCATTTTCTTTTACTT        | 135        |              |
| Db                    | 1               | TGATTCTACTGGCTGAAATAATGAATAAAGATGGATTCTTATCATTTTCTTTTACTT        | 60         |              |
| Qy                    | 136             | TTTATTGGGACTTCAGAGACACAGGTTGATGTTTTCCAAATGTCGTTTCTGGTACTAGGTCAC  | 195        |              |
| Db                    | 61              | TTTATTGGGACTTCAGAGACACAGGTAAGATGTTTTCCAAATGTCGTTTCTGGTACTAGGTCAC | 120        |              |
| Qy                    | 196             | GATATATACCAATCTCTTCAAATTTCTACAAACATACACCTCACCTGTTACTAGATAAGTCACA | 255        |              |
| Db                    | 121             | GATATATACCAATCTCTTCAAATTTCTACAAACATACACCTCACCTGTTACTAGATAAGGCGCT | 180        |              |
| Qy                    | 256             | CCAAATGTAAACAAACACAGGCGCTCCAGTCTTCTTAGCCGGGGAAAGATCGGATCTGCT     | 315        |              |
| Db                    | 181             | TCTTAATG-----AACCAGGCGCTCCAGTCTTCTTAGCCGGGGAAAGATCGGATCTGCT      | 234        |              |

QY 316 GGGATTCCTTCGTCTTGGGAATACACCACTTAATCCAAATGGGAAGGATTAATATCTTACATT 375  
DB 235 GGGATTCCTTCGTCTTGGGAATACACCACTTAATCCAAATGGGAAGGATTAATATCTTACATT 294  
QY 376 GTCAATATAAGGAAGTTGTCGGTGGATGCAAAACAGTATATACACAAGTCAGATCAAAAG 435  
DB 295 GTCAATATAAGGAAGTTGTCGGTGGATGCAAAACAGTATATACACAAGTCAGATCAAAAG 354  
QY 436 CAGACAGTCGGAAGTTCTTCTTACTTAATCTTAATCTTGGAAACACATATGAAATTAAG 495  
DB 355 CCAGACAGTCGGAAGTTCTTCTTACTTAATCTTAATCTTGGAAACACATATGAAATTAAG 414  
QY 496 GTTGTCTGTGAAACAGTCGTGCAATTTGGAGTGTGTTAGTGATCCATTTCTCTTCCAAACT 555  
DB 415 GTAGCTGTGCTGAAACAGTCGTGCAATTTGGAGTGTGTTAGTGATCCATTTCTCTTCCAAACT 474  
QY 556 GCAGAAAGTGCTCCAGGAAAGTGGTGAATCTCACAGTTTGAGGCTTACACGCTTCAGCA 615  
DB 475 GCAGAAAGTGCTCCAGGAAAGTGGTGGATTTCCACAGGTGAGGCTGTC--CCGTTCCAGCA 532  
QY 616 GTTAAGCTGATTTGGTATTTACTCTGGCAACCAATGGCAAAATTTACAGAGCTTCAAGATT 675  
DB 533 G-TAAGCTGATGTGTA--TACCCTCGGCAACCAAA-AAAAAAATTTACCAGCTTCAAGATT 588  
QY 676 AGTGTCAAAACATGCCAGAAAGTGGATAGTAGTGAAAGATGTCTCAATCAGAGTAGAGGAC 735  
DB 589 AGTGTCAAGCATTAACAGAAAGTGGATAGTAGTGAAAGAGTGTCAATCAGAGTGGAGTGC 648  
QY 736 ATTTTGACTGGGAAATTTGCC---AGAAATGCAATGAAATAGTGAATCTTTTTTATGGAGT 792  
DB 649 ATTTTAAGTGCTTCCCTTCTTGTGCACTGCAACGAGAAATAGTGAATCTTTTTTATGGAGT 708  
QY 793 AAGCCAGGCCCTTCTCCAAACCTTGTGTAGAGTTACCTCCATCGGCTACCAACATTCAC 852  
DB 709 ACAGCCAGGCCCTTCTCCAAACCTTGTGTAGAGTTTACACCTCCATCGGCTACCAACATTCAC 768  
QY 853 TCAAGCACCTTGACACAGAAATGAGATCAGCTCTGTGTGGAAGAGGCTATCAGTTTGTGA 912  
DB 769 TCAGCACGCTTGACACAGAAATGAGATCAGCTCTGT---GAAAGAGCCTATCAGTTTGTGA 825  
QY 913 GTGACACCTTGAGACCTTTATACAAATATCTTTTTTGAAGTTTCAGCTGCTACAACTGAA 972  
DB 826 GTGACACACTTGAGACCTTTATACAAATATCTTTTTTGAAGTTTCAGCTGCTACAACTGAA 885  
QY 973 GAGGTTATATTGATAGTAGATTGTACAGAACACCAAGAAATCAGTGCCTGAGGACCAACA 1032  
DB 886 GCAGGTTATATTGATAGTAGATTGTACAGAACACCAAGAAATCAGTGCCTGAGGACCAACA 945  
QY 1033 CAAAACCTGGTAAACAGCAACATCACAGGAAGTCTCTTTTCAATTTTATGGGACCCACCA 1092  
DB 946 CAAAACCTGGTAAACAGCAACATCACAGGAAGTCTCTTTTCAATTTTATGGGACCCACCA 1005  
QY 1093 ACTATAGTAACAGGGAATTTAGTTATAGAGTTGAAATATATGGACCAT---CAGGTGCG 1149  
DB 1006 ACTATAGTAACAGGGAATTTAGTTATAGAGTTGAAATATATGGACCATCAGCAGGTGCG 1065  
QY 1150 ATTTTGGATTAACAGCAAAAGACCTCAAGTTTGCATTTCACTAACCTTAACACCATTTACA 1209  
DB 1066 ATTTTGGATTAACAGCAAAAGACCTCAAGTTTGCATTTCACTAACCTTAACACCATTTACA 1125  
QY 1210 ATGTATGATGCTATATTGGGCTGAAACACAGTGCAGGAGCTGGGGCCCAAGTCAAAATATT 1269  
DB 1126 ATGTATGATGCTATATTGGGCTGAAACACAGTGCAGGAGCTGGGGCCCAAGTCAAAATATT 1185  
QY 1270 TCAGTATTCACCTCCACAGATGTTCCAGGGGCAAGTGTGTTGATTAACAATTCGAGAGGTA 1329  
DB 1186 TCAGTATTCACCTCCACAGATGTTCCAGGGGCAAGTGTGTTGATTAACAATTCGAGAGGTA 1245  
QY 1330 GAATCCACCGCAAGTAAGAAATTTACTTGGAGAAACCAACGACCAACCAATGGAATTTAATAC 1389  
DB 1246 GAATCCACCGCAAGTAAGAAATTTACTTGGAGAAACCAACGACCAACCAATGGAATTTAATAC 1305

QY 1390 CAATACCGAGTGAAGTGTCTAGTTCCAGAGACAGGAATAATTTTGGAAAAATATCTTTGCTC 1449  
DB 1306 CAATACCGAGTGAAGTGTCTAGTTCCAGAGACAGGAATAATTTTGGAAAAATATCTTTGCTC 1365  
QY 1450 ACTGGAATAATAGTATATATAATGACCCCATGGCTCCAGAAATTTGTGAAACATAGTAGAG 1509  
DB 1366 ACTGGAATAATAGT---ATAAATGACCCCATGGCTCCAGAAATTTGTGAAACATAGTAGAG 1422  
QY 1510 CCAATGGTAGGATTAATATAGGGTTTACGAGAGATGTCTGTGACCTTTCACCTCAGTTGCT 1569  
DB 1423 CCAATGGTAGGATTAATATAGGGTTTACGAGAGATGTCTGTGACCTTTCACCTCAGTTGCT 1482  
QY 1570 ACATTTATATATAACAGCCATCCAGATAAAATCTTTCTGCAAGGAATAGAGCTGAAGAC 1629  
DB 1483 ACATTTATATATAACAGCCATCCAGATAAAATCTTTCTGCAAGGAATAGAGCTGAAGAC 1542  
QY 1630 CAGACTTCAACAGTTGTAACTTAAGGAATCAGTATATATTACTGACATTTGCACTGCAACAG 1689  
DB 1543 CAGACTTCAACAGTTGTAACTTAAGGAATCAGTATATATTACTGACATTTGCACTGCAACAG 1602  
QY 1690 CTGCTCTTATGTTTATCAGGAGACTTGTACTTTCTACTGAGCACATGATTTAGTGTATCTGCT 1749  
DB 1603 CTGACTTATGTTCTTATCAGATTAAGGAGATTTTGGGCTGAGACAAATGGGGTTTCTAGA 1662  
QY 1750 TTTCCATCATCGGGAAGAGCACCAACAGTTCTCAGTGTGTAGGACACGTCAGCAAGTG 1809  
DB 1663 TATACAATCATG-----TCATCTGCAAGCAGGACAAATTTGACTTTCCCGCAGGC 1710  
QY 1810 CCAAGCTCCATTAATAATTTATAAACTATAAAATATATTAGTTCTTCATCTATTTTGTATAT 1869  
DB 1711 CCTTTGTGAGCCCAAAATTTTCCAGAGTTTACATGTTTACCATAACAGAGATATTTTACAC 1770  
QY 1870 TGGGATCTCCAGAAATATCCCAATGGAATAATACTCACTATACGATTTATGCAATGGA 1929  
DB 1771 TGGGATCTCCAGAGTCTCTGTATTTTTCATCATTAACCTTATCAGATTTTGTGATGTGAA 1830  
QY 1930 TTGGAATCAAAACAGAGACATTTCCAGATAACTTACCATAGATAACAGCTTTTCTATAACAGG 1989  
DB 1831 AACCAATCCAGAGTATTTATTTAAGGACATTTAAACAGTTTGTCTCTTCTTCTTATAGGG 1890  
QY 1990 TTAAAGAAATACACAAATATACAAATGAGAGTGGAGGCTCAACCCACGATGGAGAAAGT 2049  
DB 1891 TTAAAGAAATACACAAATATACAAATGAGAGTGGAGGCTCAACCCACGTTGGAGAAAGT 1950  
QY 2050 TCTTTGTCTGAAGAAATGACATCTTTGTGAGAACTTCAGAGATGAAACCGGAATCATCA 2109  
DB 1951 TCTTTGTCTGAAGAAATGACATCTTTGTGAGAACTTCAGAGATGAAACCGGAATCATCA 2010  
QY 2110 CCTCAAGATGTGAAAGTAAATTGATGTTTACCGCAGATGAAATAAAGTTGAAAGTGGTCA 2169  
DB 2011 CCTCAAGATGTGAAAGTAAATTGATGTTTACCGCAGATGAAATAAAGTTGAAAGTGGTCA 2070  
QY 2170 CCGGAAAGCCCAATGGGATCATTTATGCTTATGAAAGTCTATATAAAAAATATAGATACT 2229  
DB 2071 CCGGAAAGCCCAATGGGATCATTTATGCTTATGAAAGTCTATATAAAAAATATAGATACT 2130  
QY 2230 TTATATATGAAGAACACATCAACAGACATATAATTTAAGGAACTTAAAGACTCACACC 2289  
DB 2131 TTATATATGAAGAACACATCAACAGACATATAATTTAAGGAACTTAAAGACTCACACC 2190  
QY 2290 CTCTATAACATTTCTGTAAGGCTTTTACACAGATTTGGTCAATGGCAATCAGGATCTTTCT 2349  
DB 2191 CTCTATAACATTTCTGTAAGGCTTTTACACAGATTTGGTCAATGGCAATCAGGATCTTTCT 2250  
QY 2350 TTACTCTCTGTAAGGACTTTCCGAGACTGTGCTGTATGAGTGTGACACGAGAAATATCACTTAC 2409  
DB 2251 TTACTCTCTGTAAGGACTTTCCGAGACTGTGCTGTATGAGTGTGACACGAGAAATATCACTTAC 2310  
QY 2410 AAAAAATTTCTTCTGAGAGATTCAGCTTATCATTTCTTCCCTCCCAAGTAGTCCCAATGGA 2469  
DB 2311 AAAAAATTTCTTCTGAGAGATTCAGCTTATCATTTCTTCCCTCCCAAGTAGTCCCAATGGA 2370  
QY 2470 ATCATAAAAAATATATACAAATTTATCTCAAGAGAAAGTAATGGAAATGAGGAAAGAACTATA 2529

Db 2371 ATCATCAAAATATACAAATTTATCTCAAGAGAGTAATGGAAATGAGAAAGAACTATA 2430  
QY  
Db 2530 AATAACAACCTCTTTAAACCCAAACATTAAGAGTACTGAAGAAATATACCCCAATATATCAAT 2589  
QY  
Db 2431 AATAACAACCTCTTTAAACCCAAACATTAAGAGTCTGAAGAAATATACCCCAATATATCAAT 2490  
QY  
Db 2590 GAGGTCTCTGTAGTACACTGAAGAGTGAAGAGTTCGAGTCTCCCAATAATATATCTG 2649  
Db 2491 GAGGTCTCTGTAGTACACTGAAGAGTGAAGAGTTCGAGTCTCCCAATAATATATCTG 2550  
QY  
Db 2650 ACGGAGGAGAGTCTCTGATCTCTCCCTCAAGACTCTCTGTAAACAGTTCGTCTGT 2709  
QY  
Db 2551 ACGGAGGAGAGTCTCTGATCTCTCCCTCAAGACTCTCTGTAAACAGTTCGTCTGT 2610  
QY  
Db 2710 GTCAACGGTGAAGTGTCTATGCGCAACCAACCCCTGGAGCCAAATGGAAATATCTTTATTAC 2769  
QY  
Db 2611 GTCAACGGTGAAGTGTCTATGCGCAACCAACCCCTGGAGCCAAATGGAAATATCTTTATTAC 2670  
QY  
Db 2770 ACAGTTTATGTCTGGAATAGATCATCATTTAAACATTTAATGTCTCACTGAACATCATTTG 2829  
Db 2671 ACAGTTTATGTCTG --- AGATCATCATTTAAACATTTAATGTCTCACTGAACATCATTTG 2727  
QY  
Db 2830 GAGTTATCAGATTGGATTATTAATGTTGAATACAGTGTCTTATGTAAACAGTGTAGCACCAGA 2889  
Db 2728 GAGTTATCAGATTGGATTATTAATGTTGAATACAGTGTCTTATGTAAACAGTGTAGCACCAGA 2787  
QY  
Db 2890 TTTGGTGTGGGAAACAGAGCAATATCATTTAGCTTTCAAACACAGAGGAGCACCACA 2949  
Db 2788 TTTGGTGTGGGAAACAGAGCAATATCATTTAGCTTTCAAACACAGAGG --- ACCA 2844  
QY  
Db 2950 AGCGATCCCTCCAAAGATGTTTATTATGCAAACTCAGTCTCTCATCAATAATCTTTTC 3009  
Db 2845 AGCGATCCCTCCAAAGATGTTTATTATGCAAACTCAGTCTCTCATCAATAATCTTTTC 2904  
QY  
Db 3010 TGGACACCTCTCTCAAACCTTAATGGGATTTACAATAATTTACTCTGTTTATTACAGAAAT 3069  
Db 2905 TGGACACCTCTCTCAAACCTTAATGGGATTTACAATAATTTACTCTGTTTATTACAGAAAT 2964  
QY  
Db 3070 ACTTCAGGTACTTTTATGAGAAATTTTACACTCCATGAATTAACCAATGACTTTGACAAAT 3129  
Db 2965 ACTTCAGGTACTTTTATGAGAAATTTTACACTCCATGAATTAACCAATGACTTTGACAAAT 3024  
QY  
Db 3130 ATGACTGTATCCAAATATAGATAAATGACAATAATTCAGCTACTATACATTTTGGTTA 3189  
Db 3025 ATGACTGTATCCAAATATAGATAAATGACAATAATTCAGCTACTATACATTTTGGTTA 3084  
QY  
Db 3190 ACAGCAAGTACTTCAGTTGGAAATGGGAAATGAAGCAGTGACATCAATGAAGTATACACA 3249  
Db 3085 ACAGCAAGTACTTCAGTTGGAAATGGGAAATGAAGCAGTGACATCAATGAAGTATACACA 3144  
QY  
Db 3250 GATCAAGACATACCTGAAGGTTTGTGGAAACCTGACATTCGAATCCCAATTCGTCACAT 3309  
Db 3145 GATCAAGACATACCTGAAGGTTTGTGGAAACCTGACATTCGAATCCCAATTCGTCACAT 3204  
QY  
Db 3310 GCAATAAATGTAGCTGGGTCCACCGGTCCAAACCAACCGGTCTAGTCTTCTACTATGTT 3369  
Db 3205 GCAATAAATGTAGCTGGGTCCACCGGTCCAAACCAACCGGTCTAGTCTTCTACTATGTT 3264  
QY  
Db 3370 TCACTGATCTTACAGCAGACTCTCTGCCATGTGAGACCACTCTTGTGTTACATATGAGAGA 3429  
Db 3265 TCACTGATCTTACAGCAGACTCTCTGCCATGTGAGACCACTCTTGTGTTACATATGAGAGA 3324  
QY  
Db 3430 AGCATATATTTGATATCTGGAAATACACTGATTTATATTAATAAATTAATCCATCA 3489  
Db 3325 AGCATATATTTGATATCTGGAAATACACTGATTTATATTAATAAATTAATCCATCA 3384  
QY  
Db 3490 ACAGAAAGGGATCTCTGATACCTATATGCCCCAGCTATACATCAAGACTGAAGAAGAT 3549  
Db 3385 ACAGAAAGGGATCTCTGATACCTATATGCCCCAGCTATACATCAAGACTGAAGAAGAT 3444  
QY  
Db 3550 GTCCCAAGAACTTCAACCAATATCAACATTTTAAACCACTTTCTCTACCTCAGTTCTC 3609

Db 3445 ATCCCAGAAACTTTCACCAATATCAACACTTTTAAACACCTTTCTCTACCTCAGTTCTC 3504  
QY  
Db 3610 TTATCATGGGATCCCCAGTAAAGCCAAATGGTGCATAATTAAGTTATGATTTAACTTTA 3669  
QY  
Db 3505 TTATCATGGGATCCCCAGTAAAGCCAAATGGTGCATAATTAAGTTATGATTTAACTTTA 3564  
QY  
Db 3670 CNAAGACCAATGAATATTTCTTTTCATTAATCTCTGATTAATTAATATATTTGGAAGAG 3729  
Db 3565 CNAAGACCAATGAATATTTCTTTTCATTAATCTCTGATTAATTAATATATTTGGAAGAG 3624  
QY  
Db 3730 CTTTTCACCAATTTACATATATATAGCTTTTGTGCGCGCAAGAACTAGAAAAGGACTTGT 3789  
Db 3625 CTTTTCACCAATTTACATATATATAGCTTTTGTGCGCGCAAGAACTAGAAAAGGACTTGT 3684  
QY  
Db 3790 CTTTTCACCAATTTCTTTTACACAGATGAGTCAAGTCCGCTGAGCACTCCACAAAT 3849  
Db 3685 CTTTTCACCAATTTCTTTTACACAGATGAGTCAAGTCCGCTGAGCACTCCACAAAT 3744  
QY  
Db 3850 TTGACTTTAATCAACTGTACTTTCAGACTTTTGTATGGCTGAATGGAGCCCAAGTCTCTT 3909  
Db 3745 TTGACTTTAATCAACTGTACTTTCAGACTTTTGTATGGCTGAATGGAGCCCAAGTCTCTT 3804  
QY  
Db 3910 CCAGGTGGTATTGTTTAAAGTATATAGTTTAAATTTTCAATGAACATGAACATGACACTATA 3969  
Db 3805 CCAGGTGGTATTGTTTAAAGTATATAGTTTAAATTTTCAATGAACATGAACATGACACTATA 3864  
QY  
Db 3970 TATTATAAGAAATATATCAGGATTTAAACCTGAAGCCAAACTTTGTTGGACTGGAACAGTCT 4029  
Db 3865 TATTATAAGAAATATATCAGGATTTAAACCTGAAGCCAAACTTTGTTGGACTGGAACAGTCT 3924  
QY  
Db 4030 AGCACTACTCTATCCGTGTATCTCGTTTACCAAAAGTTGGAAATGGCAATCAATTTAGT 4089  
Db 3925 AGCACTACTCTATCCGTGTATCTCGTTTACCAAAAGTTGGAAATGGCAATCAATTTAGT 3984  
QY  
Db 4090 AATGTAGTAAATTTCAACCAACCAAGATTCAGTTCCAGATGTCTGCGAGAAATGCAAGTGC 4149  
Db 3985 AATGTAGTAAATTTCAACCAACCAAGATTCAGTTCCAGATGTCTGCGAGAAATGCAAGTGC 4044  
QY  
Db 4150 ATGGCAACTAGCTGGCAGTCAAGTTTGTAGTAAATGGGATCCACCCAAAGGCAAAATGGA 4209  
Db 4045 ATGGCAACTAGCTGGCAGTCAAGTTTGTAGTAAATGGGATCCACCCAAAGGCAAAATGGA 4104  
QY  
Db 4210 ATATAAGCAGTATATGGTAAAGGAAATCTACAAAAGTTTCTCCCAAGAT 4269  
Db 4105 ATATAAGCAGTATATGGTAAAGGAAATCTACAAAAGTTTCTCCCAAGAT 4164  
QY  
Db 4270 CACATGTACACTTTCATAAAGCTTCTGCCAATACCTCATATGTCTTTAAAGTAAAGCT 4329  
Db 4165 CACATGTACACTTTCATAAAGCTTCTGCCAATACCTCATATGTCTTTAAAGTAAAGCT 4224  
QY  
Db 4330 TCAACCTCAGCTGGTGAAGGATGAAGCAATGCCATGTGAGCACTACCTGAAACA 4389  
Db 4225 TCAACCTCAGCTGGTGAAGGATGAAGCAATGCCATGTGAGCACTACCTGAAACA 4284  
QY  
Db 4390 GTTCCAGTGTCCCAAAATATCTTTTCTGATGTTCAGTCAACTAGTGAACATTTG 4449  
Db 4285 GTTCCAGTGTCCCAAAATATCTTTTCTGATGTTCAGTCAACTAGTGAACATTTG 4344  
QY  
Db 4450 ACATGATGAAGCTGACACTATCTCTGCTACTTTTCAAAATTAACAAATTAACCACTCA 4509  
Db 4345 ACATGATGAAGCTGACACTATCTCTGCTACTTTTCAAAATTAACAAATTAACCACTCA 4404  
QY  
Db 4510 CTTTCTGCTCAAAATTAACAAATTAACCAATTAACCAATTAACCAATTAACCAATTA 4569  
Db 4405 CTTTCTGCTCAAAATTAACAAATTAACCAATTAACCAATTAACCAATTAACCAATTA 4464  
QY  
Db 4570 CAATACCTCTATGAAGCTCACTTAATCTGAAGAGA CAGTATATGGAATTAAGAAATTTAGA 4629  
Db 4465 CAATACCTCTATGAAGCTCACTTAATCTGAAGAGA CAGTATATGGAATTAAGAAATTTAGA 4524  
QY  
Db 4630 TGGTATAGTTCCAAAGTGGCTGCCAGCACCAGTCTGCTATGCGCAATGCTTCAACCTGG 4689  
Db 4525 TGGTATAGTTCCAAAGTGGCTGCCAGCACCAGTCTGCTATGCGCAATGCTTCAACCTGG 4584



QY 4690 ATTTCTACAAACCTCTGCTGCTCCAGATGCTCTCTGAAATGTTTCATGTAGTA 4749  
DB 4585 ATTTCTACAAACCTCTGCTGCTCCAGATGCTCTCTGAAATGTTTCATGTAGTA 4644  
QY 4750 GCAACATCACTTTTATGATCAGCATAGCTGAGTGAACCTGCTGTCATTAATGAGCA 4809  
DB 4645 GCAACATCACTTTTATGATCAGCATAGCTGAGTGAACCTGCTGTCATTAATGAGCA 4704  
QY 4810 ACATGTTATCTGATGATCAAAATCGGTGATGATGATGAAATTTAATATATCTTCTATC 4869  
DB 4705 ACATGTTATCTGATGATCAAAATCGGTGATGATGATGAAATTTAATATATCTTCTATC 4764  
QY 4870 RAGTCAAAATGAAGAAATTAACCATAGAAATTTAAGATTTTGAATATTTCAAGGTTAT 4929  
DB 4765 RAGTCAAAATGAAGAAATTAACCATAGAAATTTAAGATTTTGAATATTTCAAGGTTAT 4824  
QY 4930 TCTGTAGTATCACTGCTATTTACTGGAACATTTAGTGTGCTATATGTAGAAGGAGTCA 4989  
DB 4825 TCTGTAGTATCACTGCTATTTACTGGAACATTTAGTGTGCTATATGTAGAAGGAGTCA 4884  
QY 4990 AGTGCTGAAATGATTGTTACTACTTTAGAAATCAGCCCCAAAGGACCCACTTAACAATG 5049  
DB 4885 AGTGCTGAAATGATTGTTACTACTTTAGAAATCAGCCCCAAAGGACCCACTTAACAATG 4944  
QY 5050 ACATTTTCAAGATACCATGATGAAATTTCAAAATTTCAATTAACGTTCTCTCTCTTCT 5109  
DB 4945 ACATTTTCAAGATACCATGATGAAATTTCAAAATTTCAATTAACGTTCTCTCTCTTCT 5004  
QY 5110 CAACTTAATGAAATATCAAGTATATCAAGTCTGTTTACCGAAGATGATCTACT 5169  
DB 5005 CAACTTAATGAAATATCAAGTATATCAAGTCTGTTTACCGAAGATGATCTACT 5064  
QY 5170 GCTGTCAGATTCACAACTCTAGTATTTATACAGAAACCAACACATTTGTCATTCATG 5229  
DB 5065 GCTGTCAGATTCACAACTCTAGTATTTATACAGAAACCAACACATTTGTCATTCATG 5124  
QY 5230 CTAGAAGGACTAAAGGTGGACATACATACATATCAGTGTGTTACGCGATCAATAGTCT 5289  
DB 5125 CTAGAAGGACTAAAGGTGGACATACATACATATCAGTGTGTTACGCGATCAATAGTCT 5184  
QY 5290 GGTGAGGCTCAAGGTTCCGATGAGAAATACCATGATATCAAGCTCCAGACAGCA 5349  
DB 5185 GGTGAGGCTCAAGGTTCCGATGAGAAATACCATGATATCAAGCTCCAGACAGCA 5244  
QY 5350 AAAACCAACCAACCCCTATTTATGATGCCACAGGAAACCTGCTTGTGATTTCAACA 5409  
DB 5245 AAAACCAACCAACCCCTATTTATGATGCCACAGGAAACCTGCTTGTGATTTCAACA 5304  
QY 5410 ATTTCAATCAGATGCAATATGTTACTACAGTATGATCATGGAACCAATTAATAATGTA 5469  
DB 5305 ATTTCAATCAGATGCAATATGTTACTACAGTATGATCATGGAACCAATTAATAATGTA 5364  
QY 5470 CAAAGTCTTGCAGCAAAACAGGACTCAGCATGATGGAATGTAACAAAGTGTATGAT 5529  
DB 5365 CAAAGTCTTGCAGCAAAACAGGACTCAGCATGATGGAATGTAACAAAGTGTATGAT 5424  
QY 5530 GCATATTTTAAAGCAAGGCCATATTTTACAAATGAAGGCTTTCTTAACCCCTCCATGT 5589  
DB 5425 GCATATTTTAAAGCAAGGCCATATTTTACAAATGAAGGCTTTCTTAACCCCTCCATGT 5484  
QY 5590 ACAGAAGGAACCAAGTTTATGTCATGGAATGAAGAAATCTACATATAGTGTCTGATAT 5649  
DB 5485 ACAGAAGGAACCAAGTTTATGTCATGGAATGAAGAAATCTACATATAGTGTCTGATAT 5544  
QY 5650 GCATGATCATTTCTTGGCAATGAAGCAAAATTTGCAATGGACCACTGAAACCAAAAG 5709  
DB 5545 GCATGATCATTTCTTGGCAATGAAGCAAAATTTGCAATGGACCACTGAAACCAAAAG 5604  
QY 5710 CAATATTTTAAATTTTATAGCTACAAATATTTATGGGCAATTTTACTGACTCTGATAT 5769  
DB 5605 CAATATTTTAAATTTTATAGCTACAAATATTTATGGGCAATTTTACTGACTCTGATAT 5664

QY 5770 TCTGACCTGTGTTAAGACTTTTAGGGGAAGACTTTTTCAGAAAGAACCGTAGAGATCATTTCT 5829  
DB 5665 TCTGACCTGTGTTAAGACTTTTAGGGGAAGACTTTTTCAGAAAGAACCGTAGAGATCATTTCT 5724  
QY 5830 TCCGTCACTTTGTGTTATCTTTCAATAAATTTCTCTTGGAAACAGCTATTTTTCGCAATTTGCA 5889  
DB 5725 TCCGTCACTTTGTGTTATCTTTCAATAAATTTCTCTTGGAAACAGCTATTTTTCGCAATTTGCA 5784  
QY 5890 AGAATTCGACAG 5949  
DB 5785 AGAATTCGACAG 5844  
QY 5950 GACACTAAATTTGAAGCTGATCAGCTCATCAGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAG 6009  
DB 5845 GACACTAAATTTGAAGCTGATCAGCTCATCAGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5904  
QY 6010 AGATTAAACCGGCCCAATTAAGCAAGAAATCTCTTCTGCAACATGTTTGAAGAGAGCTTTGCA 6069  
DB 5905 AGATTAAACCGGCCCAATTAAGCAAGAAATCTCTTCTGCAACATGTTTGAAGAGAGCTTTGCA 5964  
QY 6070 AACAAACCTTAAGTTTCAAG 6129  
DB 5965 AACAAACCTTAAGTTTCAAG 6024  
QY 6130 TCTTCAACTGATGCTGATCTGCTTTGGAAATAGAGCAAAACCGTTTCCCAACATAAAA 6189  
DB 6025 TCTTCAACTGATGCTGATCTGCTTTGGAAATAGAGCAAAACCGTTTCCCAACATAAAA 6084  
QY 6190 CCATATAATAATAAATTAAG 6249  
DB 6085 CCATATAATAATAAATTAAG 6141  
QY 6250 TATATAATGCGAGCTATATTTCTGTTTATTTATGTCAAATGAATTTTATGCTACTCAA 6309  
DB 6142 TATATAATGCGAGCTATATTTCTGTTTATTTATGTCAAATGAATTTTATGCTACTCAA 6201  
QY 6310 GGTCACTTACAGGAACAGTTGGAGATTTTGGAGAAATGTTGGAGAAACAGAGGCAAAA 6369  
DB 6202 GGTCACTTACAGGAACAGTTGGAGATTTTGGAGAAATGTTGGAGAAACAGAGGCAAAA 6261  
QY 6370 ACATTAGTAAATGCTAAACACAGTGTGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6429  
DB 6262 ACATTAGTAAATGCTAAACACAGTGTGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6321  
QY 6430 CCAGAGGACAAACAGGAG 6489  
DB 6322 CCAGAGGACAAACAGGAG 6381  
QY 6490 GATGTTCAAAATAGATTTGAGCTATCAGGATCTGAAATTTGAAGAGAGAGAGAGAGAGAGAG 6549  
DB 6382 GATGTTCAAAATAGATTTGAGCTATCAGGATCTGAAATTTGAAGAGAGAGAGAGAGAGAGAGAG 6441  
QY 6550 ACTGTTGCAAGAGTGTAACTTTTACTGCTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6609  
DB 6442 ACTGTTGCAAGAGTGTAACTTTTACTGCTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6501  
QY 6610 CCTTAATTTCACTTTGTGAGAGTGTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6669  
DB 6502 CCTTAATTTCACTTTGTGAGAGTGTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6561  
QY 6670 ATTTGTTCACTGCTGCTGAG 6729  
DB 6562 ATTTGTTCACTGCTGCTGAG 6621  
QY 6730 ACACAAACATATAAATGACCATGATTTTGTGAGATATATGAGACTAGTAGTGAACCTGAGA 6789  
DB 6622 ACACAAACATATAAATGACCATGATTTTGTGAGATATATGAGACTAGTAGTGAACCTGAGA 6681  
QY 6790 AGTGAAGAGATGTCATGCTGAG 6849  
DB 6682 AGTGAAGAGATGTCATGCTGAG 6741  
QY 6850 CTGGATCTCTTATCAATTAAGGAAGTAATCAGCCCATCTGTTTGTGTTAACTATTTCAGCA 6909



Db 6742 CTGGAATCTTATCAATAAGGAAGTAATCAGCCCATCTGTTTGTGTAATCTATTACGCA 6801  
Qy 6910 CTTTCAGAAAGATGACCTCTTTTGACGCGCCATGGAA---GGTGATGTTGAGCTTTGAATGGAA 6966  
Db 6802 CTTTCAGAAAGATGACCTCTTTTGACGCGCCATGGAAAGGTGTTGATGTTGAGCTTTGAATGGAA 6861  
Qy 6967 GAAACCACTATGTAAATATTCAGACCAAGAGATACAAATGGAAGATTTTTAAATCCCA 7026  
Db 6862 GAAACCACTATGTAAATATTCAGACCAAGAGATACAAATGGAAGATTTTTAAATCCCA 6921  
Qy 7027 GGGGCGAAAGTACCCCTCATCTTCCGAAATGAAATGCAACCTTAAAGAAATATCT 7086  
Db 6922 GGGGCGAAAGTACCCCTCATCTTCCGAAATGAAATGCAACCTTAAAGAAATATCT 6981  
Qy 7087 ATGCTTCTCTCAC 7099  
Db 6982 ATGCTTCTCTCAC 6994

## RESULT 7

US-10-218-779-5

; Sequence 5, Application US/10218779

; Publication No. US20040029222A1

; GENERAL INFORMATION:

; APPLICANT: Edinger, Shlomit

; APPLICANT: MacDougall, John

; APPLICANT: Millet, Isabelle

; APPLICANT: Ellerman, Karen

; APPLICANT: Stone, David

; APPLICANT: Gerlach, Valerie

; APPLICANT: Grosse, William

; APPLICANT: Alsobrook II, John

; APPLICANT: Lepley, Denise

; APPLICANT: Rieger, Daniel

; APPLICANT: Burges, Catherine

; APPLICANT: Casman, Stacie

; APPLICANT: Spytek, Kimberly

; APPLICANT: Boldog, Ferenc

; APPLICANT: Li, Li

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Mishra, Vishnu

; APPLICANT: Patturajan, Meera

; APPLICANT: Shenoy, Suresh

; APPLICANT: Rastelli, Luca

; APPLICANT: Tchernev, Velizar

; APPLICANT: Vernet, Corine

; APPLICANT: Zerhusen, Bryan

; APPLICANT: Malyankar, Uriel

; APPLICANT: Guo, Xiaojia

; APPLICANT: Miller, Charles

; APPLICANT: Gangoli, Baha

; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

; FILE REFERENCE: 21402-214

; CURRENT APPLICATION NUMBER: US/10/218, 779

; CURRENT FILING DATE: 2002-08-14

; PRIOR APPLICATION NUMBER: 60/253,834

; PRIOR FILING DATE: 2000-11-29

; PRIOR APPLICATION NUMBER: 60/250, -926

; PRIOR FILING DATE: 2000-11-30

; PRIOR APPLICATION NUMBER: 60/264,180

; PRIOR FILING DATE: 2001-01-25

; PRIOR APPLICATION NUMBER: 60/313,656

; PRIOR FILING DATE: 2001-08-20

; PRIOR APPLICATION NUMBER: 60/327,456

; PRIOR FILING DATE: 2001-10-05

; NUMBER OF SEQ ID NOS: 216

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 5

; LENGTH: 6994

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-218-779-5

Query Match 90.8%; Score 6453; DB 17; Length 6994;  
Best Local Similarity 96.3%; Pred. No. 0;  
Matches 6770; Conservative 0; Mismatches 215; Indels 48; Gaps 14;

Qy 76 TGAATTCCTACTGCTGAAATGTAATAAGATGGATTTCTTATCATTTTCTTTTACTT 135  
Db 1 TGAATTCCTACTGCTGAAATGTAATAAGATGGATTTCTTATCATTTTCTTTTACTT 60

Qy 136 TTTATTTGGGACTTTTCAGAGACACAGGTTTGAATGTTTCCAATGTCGTTCTCGTACTAGTAC 195  
Db 61 TTTATTTGGGACTTTTCAGAGACACAGGTTTGAATGTTTCCAATGTCGTTCTCGTACTAGTAC 120

Qy 196 GATATAACCATCTCTTCAATTTCTACAACATACACCTCACCTGTTTACTAGATAGTAC 255  
Db 121 GATATAACCATCTCTTCAATTTCTACAACATACACCTCACCTGTTTACTAGATAGTAC 180

Qy 256 CCAATGTAACAAACAGGCTTCCAGTCTTCTAGCGGGGAAAGATCGGATCTGCT 315  
Db 181 TCTAATG-----AACAGGCTTCCAGTCTTCTAGCGGGGAAAGATCGGATCTGCT 234

Qy 316 GGGATTTCTTCTGTTTGGAAATACACCACTTAATCCAAATGGAAGGATTTATCTTACAT 375  
Db 235 GGGATTTCTTCTGTTTGGAAATACACCACTTAATCCAAATGGAAGGATTTATCTTACAT 294

Qy 376 GTCAAAATATAAGAAAGTTTGTCCGTGGATGCAAAAGATATATACAAAGTCAGATCAAG 435  
Db 295 GTCAAAATATAAGAAAGTTTGTCCGTGGATGCAAAAGATATATACAAAGTCAGATCAAG 354

Qy 436 CCAGACAGTCTGGAGTTCTTCTTACTTAATCTTAATCTTCTGGACCAATATGAATTAAG 495  
Db 355 CCAGACAGTCTGGAGTTCTTCTTACTTAATCTTCTGGACCAATATGAATTAAG 414

Qy 496 GTTGTCTGCTGAAACAGTCTGGCATTTGGAGTGTGTAGTGCATCTTCTTCTTCCAACT 555  
Db 415 GTAGTCTGCTGAAACAGTCTGGCATTTGGAGTGTGTAGTGCATCTTCTTCTTCCAACT 474

Qy 556 GCAGAAAGTCTCCAGGAAAGTGGTGAATCTCACAGTTTGAGGCTTCAACGCTTCAGCA 615  
Db 475 GCAGAAAGTCTCCAGGAAAGTGGTGGATTTTCCAGAGTGGAGCTGTCTCCGTTACGA 532

Qy 616 GTTAACTGATTTGGTATTTTACTCTGGCAACCAATGGCAAAATTTACGAGTTCAGAT 675  
Db 533 G-TAAGCTGATGTGGTA--TACCTCGGCAACCA--AAAAAAATTTACGAGTTCAGAT 588

Qy 676 AGTGTCAAAATGCTCCAGAGTGGGATAGTAGTGAAGATGTCTCAATCAGAGTAGAGGAC 735  
Db 589 AGTGTCAAGCATTAACAGAAAGTGGGATAGTAGTGAAGAAAGTGTCAATCAGAGTGGAGTGC 648

Qy 736 ATTTTGAATGGGAAATTTGCC---AGAAATGCAATGAGAAATAGTGAATCTTTTTTATGGAGT 792  
Db 649 ATTTTGAATGCTTCCCTTCTTCTTGGCACTGCAACGAGAAATAGTGAATCTTTTTTATGGAGT 708

Qy 793 ACAGCCAGGCTTCTCCAACTTGTAGATGTTACCTCCATCGGCTGACCAACATTTCA 852  
Db 709 ACAGCCAGGCTTCTCCAACTTGTAGATGTTACCTCCATCGGCTGACCAACATTTCA 768

Qy 853 TCAAGCAGCTTCACAGAAATGAGATCAGTCTGTGTGAAAGAGCTTATCAGTTTGTGA 912  
Db 769 TCAAGCAGCTTCACAGAAATGAGATCAGTCTGTGTGAAAGAGCTTATCAGTTTGTGA 825

Qy 913 GTGACACACTTTGAGACCTTTATACAATATCTTTTTTGAAGTTTTCAGTCTGCTCAACTGAA 972  
Db 826 GTGACACACTTTGAGACCTTTATACAATATCTTTTTTGAAGTTTTCAGTCTGCTCAACTGAA 885

Qy 973 GCAGGTTATATTGATAGTACGATTTGTACAGAACCAAGAAATCAGTGTGCTGAGGACCA 1032  
Db 886 GCAGGTTATATTGATAGTACGATTTGTACAGAACCAAGAAATCAGTGTGCTGAGGACCA 945

Qy 1033 CAAAATGCTGTAACAGGCAACATCAGGAAAGTCTTCTTCAATTTTATGGGACCCACCA 1092  
Db 946 CAAAATGCTGTAACAGGCAACATCAGGAAAGTCTTCTTCAATTTTATGGGACCCACCA 1005

QY 1093 ACTATAGTAACAGGGAAATTTAGTTATAGAGTTGAATATATGACCAT---CAGGTCCG 1149  
DB 1006 ACTATAGTAACAGGGAAATTTAGTTATAGAGTTGAATATATGACCATCAGCAGGTCCG 1065  
QY 1150 ATTTTGGATTAACAGCAAAAAGACCTCAAGTTTGGATTTGCATTTCACTAACTTAACACCATTTACA 1209  
DB 1066 ATTTTGGATTAACAGCAAAAAGACCTCAAGTTTGGATTTGCATTTCACTAACTTAACACCATTTACA 1125  
QY 1210 ATGTATGATGCTATATTTGCGGCTGAAACACAGTGCAGGGACCTGGGCCCAGGTCAATATTT 1269  
DB 1126 ATGTATGATGCTATATTTGCGGCTGAAACACAGTGCAGGGACCTGGGCCCAGGTCAATATTT 1185  
QY 1270 TCAGTATTCACCTCCACAGATGTTCCAGGGGAGTGTTTGATTTTAACTTGCAGAGGTA 1329  
DB 1186 TCAGTATTCACCTCCACAGATGTTCCAGGGGAGTGTTTGATTTTAACTTGCAGAGGTA 1245  
QY 1330 GAATCCACGCAAGTAGAATTTACTTTGGAGAAACCAAGCAACCAAAATGGAATTTAATTAAC 1389  
DB 1246 GAATCCACGCAAGTAGAATTTACTTTGGAGAAACCAAGCAACCAAAATGGAATTTAATTAAC 1305  
QY 1390 CAATACCGAGTGAAGTGTAGTTCCAGAGACAGGAATTAATTTTGGAGAAATTTTGCCTC 1449  
DB 1306 CAATACCGAGTGAAGTGTAGTTCCAGAGACAGGAATTAATTTTGGAGAAATTTTGCCTC 1365  
QY 1450 ACTGGAAATAATGAGTATATAAATGACCCCATGGCTCCAGAAATTTGTGAACATAGTAGAG 1509  
DB 1366 ACTGGAAATAATGAG---ATAATGACCCCATGGCTCCAGAAATTTGTGAACATAGTAGAG 1422  
QY 1510 CCAATGGTAGGATTAATGAGGGTTCCAGAGAGATGCTGTGACCTTCACTCACTTGTCT 1569  
DB 1423 CCAATGGTAGGATTAATGAGGGTTCCAGAGAGATGCTGTGACCTTCACTCACTTGTCT 1482  
QY 1570 ACATTTATATATAACAGCCATCCAGATAAAATTTCTTCCAGAGGAATAGCTGAGAC 1629  
DB 1483 ACATTTATATATAACAGCCATCCAGATAAAATTTCTTCCAGAGGAATAGCTGAGAC 1542  
QY 1630 CAGACTTCCACGAGTTGTAACTACAAGGAATCAGTATATTAATCTGACATTCAGCTGAACAG 1689  
DB 1543 CAGACTTCCACGAGTTGTAACTACAAGGAATCAGTATATTAATCTGACATTCAGCTGAACAG 1602  
QY 1690 CTGCTTATGTTATCAGGAGACTTGTACCTTTACCTGAGACACATGATTAATGATGCTGCT 1749  
DB 1603 CTGACTTATGTTCTTATCAGATTAAGAGATTTTGGGCTGAGACAAATTTGGGTTTCTAGA 1662  
QY 1750 TTCCACATCATGGGAGAGGACCAACCAAGTCTCAGTGTAGGACACGTCAGCAAGTG 1809  
DB 1663 TATACATCATG-----TCATCTGCAAGCAGGGAACAATTTGACTCTCCCGAGGC 1710  
QY 1810 CCAAGCTCCATTAATAAATAATAAATAATTAATTTCTTCTATCTATTTTGTATAT 1869  
DB 1711 CTTTGTGAGCCCAAAATTCAGAGTTACATGTTACATTAACAGAAATTTTATACAC 1770  
QY 1870 TGGGATCTCCAGAAATPATCCCAATGGAAATAAATCACTATACGATTTTATGCAATGGAA 1929  
DB 1771 TGGGATCTCCAGATCTGATTTTTCATCTATTAATCTTATCATTATTTTGGATGTTGAA 1830  
QY 1930 TTGGATACAAACAGACATTCAGATAAATCAACATAGATAACAGCTTCTCATACAGGG 1989  
DB 1831 AACCAATCCAAGAGTATTTATTTAAGGACATTAACACAGTTTGTCTCTGCTTATAGGG 1890  
QY 1990 TTAAGAAATACACAAAATACAAAATGAGAGTGGCAGCTCAACCCAGCATGAGAAAGT 2049  
DB 1891 TTAAGAAATACACAAAATACAAAATGAGAGTGGCAGCTCAACCCAGCTTGGAGAAAGT 1950  
QY 2050 TCTTTGTCTGAAGAAATGACATCTTTTGTGAGAACTTCAGAGATGAACCGGAATCATCA 2109  
DB 1951 TCTTTGTCTGAAGAAATGACATCTTTTGTGAGAACTTCAGAGATGAACCGGAATCATCA 2010  
QY 2110 CCTCAAGATGTCGAAGTAATGATGTTACCGCAGATGAATTAAGTTGAAGTGGTCACCA 2169  
DB 2011 CCTCAAGATGTCGAAGTAATGATGTTACCGCAGATGAATTAAGTTGAAGTGGTCACCA 2070  
QY 2170 CCCGAAAGCCCAATGGGATCATTTATGCTTATGAAAGTGTATATAAAAAATATAGATACT 2229

DB 2071 CCCGAAAGCCCAATGGGATCATTTATGCTTATGAAGTGTATATAAAAAATATAGATACT 2130  
QY 2230 TTATATATGAAGAAACACATCAACACAGACATAATATTAAAGNACTTTAAGACCTCACACC 2289  
DB 2131 TTATATATGAAGAAACACATCAACACAGACATAATATTAAAGNACTTTAAGACCTCACACC 2190  
QY 2290 CTCTATAACATTTCTGTAAAGTCTTTACACAGATTTGGTCAATGGCAATCAGGTATCTTCT 2349  
DB 2191 CTCTATAACATTTCTGTAAAGTCTTTACACAGATTTGGTCAATGGCAATCAGGTATCTTCT 2250  
QY 2350 TTACTCTCTGTAAAGACTTTGGAGACTGTGCTGTATGTCACAGAAAATATCACTTAC 2409  
DB 2251 TTACTCTCTGTAAAGACTTTGGAGACTGTGCTGTATGTCACAGAAAATATCACTTAC 2310  
QY 2410 AAAAATATTTCTTCTGGAGAGATTTGAGCTATCATTTCTTCCCCCAAGTAGTCCCAATGGA 2469  
DB 2311 AAAAATATTTCTTCTGGAGAGATTTGAGCTATCATTTCTTCCCCCAAGTAGTCCCAATGGA 2370  
QY 2470 ATCATATAAAAAATATACAAATTTATCTCAAGAGAAATTAATGGAAATGAGGAAAGAACTATA 2529  
DB 2371 ATCATATAAAAAATATACAAATTTATCTCAAGAGAAATTAATGGAAATGAGGAAAGAACTATA 2430  
QY 2530 AATACAAACCTCTTTAAACCCAAACCATTTAAAGTACTGAAGAAATATATACCAATATATCAT 2589  
DB 2431 AATACAAACCTCTTTAAACCCAAACCATTTAAAGTCTGAAGAAATATATACCAATATATCAT 2490  
QY 2590 GAGGTGTCTGTAGTACACTGAAAGGTGAAGAGTTCGAGGTGCTCCCAATAGTATACATG 2649  
DB 2491 GAGGTGTCTGTAGTACACTGAAAGGTGAAGAGTTCGAGGTGCTCCCAATAGTATACATG 2550  
QY 2650 ACGGAGGAGAGTGTCTGATTTCTCCCTCAAGACTTCTCTGTAAACAGTGTGCTGTG 2709  
DB 2551 ACGGAGGAGAGTGTCTGATTTCTCCCTCAAGACTTCTCCCTCAAGACTTCTCTGTAAACAGTGTGCTGTG 2610  
QY 2710 GTACGAGTGAAGTGTCTGATTTCTGATGCAACCAACCCCTGGAGCCAAATTTAATTTATTAC 2769  
DB 2611 GTACGAGTGAAGTGTCTGATGCAACCAACCCCTGGAGCCAAATTTAATTTATTAC 2670  
QY 2770 ACAGTTTATGCTCGAATAGATCATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2829  
DB 2671 ACAGTTTATGCTGCG---AGATCATCATTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 2727  
QY 2830 GAGTTATCAGATTTGAGATTAATTTGATTAATCAGTGTCTTATGTAACAGTACAGTACAGCAGA 2889  
DB 2728 GAGTTATCAGATTTGAGATTAATTTGATTAATCAGTGTCTTATGTAACAGTACAGTACAGCAGA 2787  
QY 2890 TTTGGTGAATGGAAACAGAAAGCAATATCATTTAGCTTTTCAACACAGAGGGAGCACA 2949  
DB 2788 TTTGGTGAATGGAAACAGAAAGCAATATCATTTAGCTTTTCAACACAGAGGG---ACCA 2844  
QY 2950 AGCGATCTCCCAAGATGTTTATGCAAACTCAGTCTTCTCATCAATATTTCTTTTC 3009  
DB 2845 AGCGATCTCCCAAGATGTTTATGCAAACTCAGTCTTCTCATCAATATTTCTTTTC 2904  
QY 3010 TCGACACCTCTTCAAACTTAATGGAATTTATACAAATTTACTCTGTTTATACAGAAAT 3069  
DB 2905 TCGACACCTCTTCAAACTTAATGGAATTTATACAAATTTACTCTGTTTATACAGAAAT 2964  
QY 3070 ACTTCAGGTATCTTTTATGCAAAATTTTACACTCCATGAACTAAACCAATGACTTTGCAAT 3129  
DB 2965 ACTTCAGGTATCTTTTATGCAAAATTTTACACTCCATGAACTAAACCAATGACTTTGCAAT 3024  
QY 3130 ATGACTGTATCCAAATTTATAGATAAATTTAGCAATTTTACGCTACTATCATTTTGTGTTA 3189  
DB 3025 ATGACTGTATCCAAATTTATAGATAAATTTAGCAATTTTACGCTACTATCATTTTGTGTTA 3084  
QY 3190 ACAGCAAGTACTTTCAGTTGGAATGGAATAAAGCAGTGACATCAATGCAAGTATACACA 3249  
DB 3085 ACAGCAAGTACTTTCAGTTGGAATGGAATAAAGCAGTGACATCAATGCAAGTATACACA 3144  
QY 3250 GATCAAGACATACCTGAAAGGTTTGTGGAACCTTGAACTTACGAATTTGCTGCTCAACT 3309

Db 3145 GATCAAGAGCTACCTGAAGGTTTGTGGAAACCTGACTTACGAATCCAAATTCCTGCAACT 3204  
Qy 3310 GCAATAAATGTAAGCTGGCTCCACCGCTCAACCAACGGTCTAGTCTTCTTACTATGTT 3369  
Db 3205 GCAATAAATGTAAGCTGGCTCCACCGCTCAACCAACGGTCTAGTCTTCTTACTATGTT 3264  
Qy 3370 TCACGTGATCTTACAGAGACTCTCTGCCATGTGAGACCACTCTTGTGTACATATGAGAGA 3429  
Db 3265 TCACGTGATCTTACAGAGACTCTCTGCCATGTGAGACCACTCTTGTGTACATATGAGAGA 3324  
Qy 3430 AGCATATATTTTGTAAATCTGGAAATAACACCTGATATATATAAATAATTAATCTCCATCA 3489  
Db 3325 AGCATATATTTTGTAAATCTGGAAATAACACCTGATATATATAAATAATTAATCTCCATCA 3384  
Qy 3490 ACAGAAAAGGGATTCTCTGATACCTATATCTGCCAGCTATACATCAAGACTGAAGAAGAT 3549  
Db 3385 ACAGAAAAGGGATTCTCTGATACCTATATCTGCCAGCTATACATCAAGACTGAAGAAGAT 3444  
Qy 3550 GTCCAGAAACTTACCAATATCAACCTTTTAAACCTTTCTCTACCTCAGTTCTC 3609  
Db 3445 ATCCAGAAACTTACCAATATCAACCTTTTAAACCTTTCTCTACCTCAGTTCTC 3504  
Qy 3610 TTATCATGGATCCCGGATGAAGCAATGCTGCAATAAAGTTATGATTTAACTTTA 3669  
Db 3505 TTATCATGGATCCCGGATGAAGCAATGCTGCAATAAAGTTATGATTTAACTTTA 3564  
Qy 3670 CAAGACCAAAATGAATAATTTCTTTCAATTAATCTGATAATTAACATAATTTGAAGAG 3729  
Db 3565 CAAGACCAAAATGAATAATTTCTTTCAATTAATCTGATAATTAACATAATTTGAAGAG 3624  
Qy 3730 CTTTCAACATTAATATATAGCTTTTGTGCGCAAGAACTAGAAAAGCACTTGGT 3789  
Db 3625 CTTTCAACATTAATATATAGCTTTTGTGCGCAAGAACTAGAAAAGCACTTGGT 3684  
Qy 3790 CTTTCCAGTATCTTTTCTTTTACACAGTAGTCACTGCGTGGACCTCCCAAAAT 3849  
Db 3685 CTTTCCAGTATCTTTTCTTTTACACAGTAGTCACTGCGTGGACCTCCCAAAAT 3744  
Qy 3850 TTGACTTTAATCAACTGACTTTCAGACTTTGTATGGCTGAATGGAGCCCAAGTCTCTT 3909  
Db 3745 TTGACTTTAATCAACTGACTTTCAGACTTTGTATGGCTGAATGGAGCCCAAGTCTCTT 3804  
Qy 3910 CCAGTGGTATGTTAAAGTATATAGTTTAAATTCATGAACATGAACCTGACACTATA 3969  
Db 3805 CCAGTGGTATGTTAAAGTATATAGTTTAAATTCATGAACATGAACCTGACACTATA 3864  
Qy 3970 TATTATAGAATATATCAGGATTTAAACTGAAGCAAACTTTGTTGGAGTGAACCACTC 4029  
Db 3865 TATTATAGAATATATCAGGATTTAAACTGAAGCAAACTTTGTTGGAGTGAACCACTC 3924  
Qy 4030 AGCCTTACTCTATCCGTGATCTGCTTTCACAAAGTTGGAATGGCAATCAATTTAGT 4089  
Db 3925 AGCCTTACTCTATCCGTGATCTGCTTTCACAAAGTTGGAATGGCAATCAATTTAGT 3984  
Qy 4090 AATGTAGTAAATTCACAAACCAAGATCAGTTCCAGATGTCGTGCAGATATGCAAGTGC 4149  
Db 3985 AATGTAGTAAATTCACAAACCAAGATCAGTTCCAGATGTCGTGCAGATATGCAAGTGC 4044  
Qy 4150 ATGGCAACTAGCTGGCAGTCACTTTTGTAGTGAATGGATCCACCCAAAAGGCAATGGA 4209  
Db 4045 ATGGCAACTAGCTGGCAGTCACTTTTGTAGTGAATGGATCCACCCAAAAGGCAATGGA 4104  
Qy 4210 ATAATAACGCAATATATGTAACAGTTGAAGGAATTTCTACAAAAGTTTCTCCCAAGAT 4269  
Db 4105 ATAATAACGCAATATATGTAACAGTTGAAGGAATTTCTACAAAAGTTTCTCCCAAGAT 4164  
Qy 4270 CACATGTACACTTTCAATAAGCTTCTTGCCAAATACCTCATATGTCCTTTAAAGTAAGACT 4329  
Db 4165 CACATGTACACTTTCAATAAGCTTCTTGCCAAATACCTCATATGTCCTTTAAAGTAAGACT 4224  
Qy 4330 TCAACTCAGCTGGTGAAGGTGATGAAGGACATGCGATGTGAGCACTACCTGAACA 4389  
Db 4225 TCAACTCAGCTGGTGAAGGTGATGAAGGACATGCGATGTGAGCACTACCTGAACA 4284

Qy 4390 GTTCCAGTGTTCACAAATATTTGCTTTTCTGTATGTTCACTCAACTAGTCAACATGTG 4449  
Db 4285 GTTCCAGTGTTCACAAATATTTGCTTTTCTGTATGTTCACTCAACTAGTCAACATGTG 4344  
Qy 4450 ACATGGTAAGACCTGACACTATCTTGGCTACTTTCAAAATTAACAAATTAACCACTCA 4509  
Db 4345 ACATGGTAAGACCTGACACTATCTTGGCTACTTTCAAAATTAACAAATTAACCACTCA 4404  
Qy 4510 CTTCTGTCTCAAAATGCAAAAGATGGGAATCCGAAGATGTTGTTGAATATCAAAAAAT 4569  
Db 4405 CTTCTGTCTCAAAATGCAAAAGATGGGAATCCGAAGATGTTGTTGAATATCAAAAAAT 4464  
Qy 4570 CAATACCTCTATGAAGCTCACTTAATCTGAAGAGACAGTATATGGAATTAAGAAAAAT 4629  
Db 4465 CAATACCTCTATGAAGCTCACTTAATCTGAAGAGACAGTATATGGAATTAAGAAAAAT 4524  
Qy 4630 TGGTATAGATCCCAAGTGGCTGCCAGCACCAATGCTGGCTATGGCAATGCTTCAAACTGG 4689  
Db 4525 TGGTATAGATCCCAAGTGGCTGCCAGCACCAATGCTGGCTATGGCAATGCTTCAAACTGG 4584  
Qy 4690 ATTTCTACAAAACTCTGCTGGCCCTCCAGATGCTCTCTGAAAAATGTTCAATGTAGTA 4749  
Db 4585 ATTTCTACAAAACTCTGCTGGCCCTCCAGATGCTCTCTGAAAAATGTTCAATGTAGTA 4644  
Qy 4750 GCAACATCACCTTTTAGCATCAGCATAAAGCTGGAGTGAACCTGCTGTCAATTAAGTGAACA 4809  
Db 4645 GCAACATCACCTTTTAGCATCAGCATAAAGCTGGAGTGAACCTGCTGTCAATTAAGTGAACA 4704  
Qy 4810 ACATGTTATCTGATGATGTCMAATCGGTAGTATATGATGAATTAATATATCTCTCATC 4869  
Db 4705 ACATGTTATCTGATGATGTCMAATCGGTAGTATATGATGAATTAATATATCTCTCATC 4764  
Qy 4870 AAGTCAAAATGAAGAAAAATAAAACCATAGAAATTTAAGATTTTGAATAATTTCAAGAGTAT 4929  
Db 4765 AAGTCAAAATGAAGAAAAATAAAACCATAGAAATTTAAGATTTTGAATAATTTCAAGAGTAT 4824  
Qy 4930 TCTGTAGTGAATCACTGCAATTTTCTGGAACATTTAGTGTGTCATATGTAGAAGGAGTCA 4989  
Db 4825 TCTGTAGTGAATCACTGCAATTTTCTGGAACATTTAGTGTGTCATATGTAGAAGGAGTCA 4884  
Qy 4990 AGTGTGAAATGATGTTTACTTACTTTAGATCAGCCCCAAGGACCCACCTTAACAACTG 5049  
Db 4885 AGTGTGAAATGATGTTTACTTACTTTAGATCAGCCCCAAGGACCCACCTTAACAACTG 4944  
Qy 5050 ACATTTCAAGAGATACAGAGTGAAGTTACAAAATTTCAATTAACCTTCTCTCTCTCT 5109  
Db 4945 ACATTTCAAGAGATACAGAGTGAAGTTACAAAATTTCAATTAACCTTCTCTCTCTCTCT 5004  
Qy 5110 CAACCTAATGGAAATATCCAAAGTATATCAAGCTCTGGTTTACCGAGAGAGATGATCCTACT 5169  
Db 5005 CAACCTAATGGAAATATCCAAAGTATATCAAGCTCTGGTTTACCGAGAGAGATGATCCTACT 5064  
Qy 5170 GCTGTCCAGATTCACAACTCAGATTTATACGAAAAACCAACACATTTGTGTCAATGCAATG 5229  
Db 5065 GCTGTCCAGATTCACAACTCAGATTTATACGAAAAACCAACACATTTGTGTCAATGCAATG 5124  
Qy 5230 CTAGAAGCACTAAAGGTGGACATACATCAATATCAGTGTGTTACGCACTCAATAGTGT 5289  
Db 5125 CTAGAAGCACTAAAGGTGGACATACATCAATATCAGTGTGTTACGCACTCAATAGTGT 5184  
Qy 5290 GGTGAGGTCCAAAGGTTCCGATGAGAAATAACCATGGATATCAAAAGCTCCAGCAGCA 5349  
Db 5185 GGTGAGGTCCAAAGGTTCCGATGAGAAATAACCATGGATATCAAAAGCTCCAGCAGCA 5244  
Qy 5350 AAAACCAAAACCAACCCCTATTTATGATGCCACAGAAAACTGCTGTGCACTTCAACA 5409  
Db 5245 AAAACCAAAACCAACCCCTATTTATGATGCCACAGAAAACTGCTGTGCACTTCAACA 5304  
Qy 5410 ATTACAAATCAGATGCCCAATATGTTACTTACAGTGTATGATGAGCACTTCAAAAAATGTA 5469  
Db 5305 ATTACAAATCAGATGCCCAATATGTTACTTACAGTGTATGATGAGCACTTCAAAAAATGTA 5364

QY 5470 CAAGTGTTCGACAGAAACAGGAGCTCAGCATGATGGAATGTAAACAAAGTGGTATGAT 5529  
Db 5365 CAAGTGTTCGACAGAAACAGGAGCTCAGCATGATGGAATGTAAACAAAGTGGTATGAT 5424  
QY 5530 GCATATTTTAAATAAGCAAGGCCATATTTTAAATAAGGCTTTTCTTAAACCTCCATGT 5589  
Db 5425 GCATATTTTAAATAAGCAAGGCCATATTTTAAATAAGGCTTTTCTTAAACCTCCATGT 5484  
QY 5590 ACAGAGAAAGACAAAGCTTTAGTGGCAATGAGGAATCTACATCATAGTGGCTGATAT 5649  
Db 5485 ACAGAGAAAGACAAAGCTTTAGTGGCAATGAGGAATCTACATCATAGTGGCTGATAT 5544  
QY 5650 GCATGCATGATTTCTGGCAATGAGCAAAATTTGCAATGGACCACTGAAACCAAAAAG 5709  
Db 5545 GCATGCATGATTTCTGGCAATGAGCAAAATTTGCAATGGACCACTGAAACCAAAAAG 5604  
QY 5710 CAATACTTATTTAAATTTAGAGCTACAAATATTTATGGGACAAATTTTACCTCGATAT 5769  
Db 5605 CAATACTTATTTAAATTTAGAGCTACAAATATTTATGGGACAAATTTTACCTCGATAT 5664  
QY 5770 TCTGACCTGTTTAGACTTTAGGGGAAGGACTTTTTCAGAAAGAACCGTAGAGATCATTTCT 5829  
Db 5665 TCTGACCTGTTTAGACTTTAGGGGAAGGACTTTTTCAGAAAGAACCGTAGAGATCATTTCT 5724  
QY 5830 TCCGTCACTTTGTGTATCCTTTCAATAATTTCTCTTGGAAACAGCTATTTTTCGCAATTTGCA 5889  
Db 5725 TCCGTCACTTTGTGTATCCTTTCAATAATTTCTCTTGGAAACAGCTATTTTTCGCAATTTGCA 5784  
QY 5890 AGAATTCGACAGAGACAGAAAGAGTGGCAATATCTCTCTAGGATGCGAGAAATTTAT 5949  
Db 5785 AGAATTCGACAGAGACAGAAAGAGTGGCAATATCTCTCTAGGATGCGAGAAATTTAT 5844  
QY 5950 GACACTAAATTTAGAGCTGATCAGCTCATCAGTGGCAGACCTGGAATGAGGACGAG 6009  
Db 5845 GACACTAAATTTAGAGCTGATCAGCTCATCAGTGGCAGACCTGGAATGAGGACGAG 5904  
QY 6010 AGATTAAACGGCGCAATAAGCAAGAAATCTCTTCTGCAACATGTTGGAAGAGCTTGCACA 6069  
Db 5905 AGATTAAACGGCGCAATAAGCAAGAAATCTCTTCTGCAACATGTTGGAAGAGCTTGCACA 5964  
QY 6070 AACAAACACCTTAAAGTTTCAAGAGAAATTTTCGGAAATTTCCAAATTTCTTTCAGGATCTT 6129  
Db 5965 AACAAACACCTTAAAGTTTCAAGAGAAATTTTCGGAAATTTCCAAATTTCTTTCAGGATCTT 6024  
QY 6130 TCTTCAACTGATGCTGATCTGCTGGAATAGAGCAAAACCGTTTCCCAACATAAAA 6189  
Db 6025 TCTTCAACTGATGCTGATCTGCTGGAATAGAGCAAAACCGTTTCCCAACATAAAA 6084  
QY 6190 CCATATAATAATAACAGAGTAAGCTGATAGCTGACGCTAGTGTTCACAGGTTCCGAT 6249  
Db 6085 CCAT---ATAATAATAACAGAGTAAGCTGATAGCTGACGCTAGTGTTCACAGGTTCCGAT 6141  
QY 6250 TATATTAATGCCAGCTATATTTCTGGTATTTATGTCCAAATGAATTTATTTGCTACTCAA 6309  
Db 6142 TATATTAATGCCAGCTATATTTCTGGTATTTATGTCCAAATGAATTTATTTGCTACTCAA 6201  
QY 6310 GGTCCACTACCAAGNACAGTTGGAGATTTTGGAGAAATGGTGGGAACACAGGGCAAAA 6369  
Db 6202 GGTCCACTACCAAGNACAGTTGGAGATTTTGGAGAAATGGTGGGAACACAGGGCAAAA 6261  
QY 6370 ACATTAGTAATGCTAAACAGTGTTTTGAAAAGGACGATCAGATGCCATCAGTATTGG 6429  
Db 6262 ACATTAGTAATGCTAAACAGTGTTTTGAAAAGGACGATCAGATGCCATCAGTATTGG 6321  
QY 6430 CCAGAGGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGATTTACAAAGCTAATGGAG 6489  
Db 6322 CCAGAGGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGATTTACAAAGCTAATGGAG 6381  
QY 6490 GATGTTCAAAATAGATTGCACTATCAGGATCTGAAATTTGAAAGCATGGGGATGGCATG 6549  
Db 6382 GATGTTCAAAATAGATTGCACTATCAGGATCTGAAATTTGAAAGCATGGGGATGGCATG 6441  
QY 6550 ACTGTTTCAGACAGTGTAACTTTTACTGCTGGCCAGAGCATGGGGTTCCTGAGAAACAGCGCC 6609

Db 6442 ACTGTTTCAGACAGTGTAACTTTTACTGCTGGCCAGAGCATGGGGTTCCTGAGAAACAGCGCC 6501  
QY 6610 CCTCTAATTTCACTTTTGTGAAGTTGGTTCGACCAAGCAGGCGACATGACACACACCTATG 6669  
Db 6502 CCTCTAATTTCACTTTTGTGAAGTTGGTTCGACCAAGCAGGCGACATGACACACACCTATG 6561  
QY 6670 ATTGTTCACTGCACTGCTGGAGTTTGGAAAGAACTGGAGTTTTTATTTGCTCTCGACCATTTA 6729  
Db 6562 ATTGTTCACTGCACTGCTGGAGTTTGGAAAGAACTGGAGTTTTTATTTGCTCTCGACCATTTA 6621  
QY 6730 ACACAACATATAAATGACCATGATTTTGTGGATATATATGGACTAGTAGCTGAACTGAGA 6789  
Db 6622 ACACAACATATAAATGACCATGATTTTGTGGATATATATGGACTAGTAGCTGAACTGAGA 6681  
QY 6790 AGTGAAGAATGTCATGCTGGCAGAACTGGCAGACAGTATATCTTTTTACACAGTGCATT 6849  
Db 6682 AGTGAAGAATGTCATGCTGGCAGAACTGGCAGACAGTATATCTTTTTACACAGTGCATT 6741  
QY 6850 CTGGATCTCTTATCAAAATTAAGGGAAGTAATCAGGCCCATCTGTTTGTAACTATTTCAGCA 6909  
Db 6742 CTGGATCTCTTATCAAAATTAAGGGAAGTAATCAGGCCCATCTGTTTGTAACTATTTCAGCA 6801  
QY 6910 CTTCAGAAGATGGAATCTTTTGGACGCCATGGAA---GGTGAATGTGAGCTTGAATGGAA 6966  
Db 6802 CTTCAGAAGATGGAATCTTTTGGACGCCATGGAAAGGTGGTGAATGTGAGCTTGAATGGAA 6861  
QY 6967 GAAACCATATGTAATAATTTACAGACCAAGAGTAACAAATGGAAGAGATTTTAAATCCCA 7026  
Db 6862 GAAACCATATGTAATAATTTACAGACCAAGAGTAACAAATGGAAGAGATTTTAAATCCCA 6921  
QY 7027 GGGGCCAAAGTTACCCCTCATCTTCCGAATTCGAATGCAACCTTAAAGAAATATCT 7086  
Db 6922 GGGGCCAAAGTTACCCCTCATCTTCCGAATTCGAATGCAACCTTAAAGAAATATCT 6981  
QY 7087 ATGCTTCTCTCAC 7099  
Db 6982 ATGCTTCTCTCAC 6994

## RESULT 8

US-10-314-232-21  
; Sequence 21, Application US/10314232  
; Publication No. US20030138932A1  
; GENERAL INFORMATION:  
; APPLICANT: MOLLER, NIELS P.H.  
; APPLICANT: MOLLER, KARIN B.  
; APPLICANT: ULLRICH, AXEL  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE  
; FILE REFERENCE: 038602/0686  
; CURRENT APPLICATION NUMBER: US/10/314,232  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/361,096  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: 08/449,609  
; PRIOR FILING DATE: 1995-05-24  
; PRIOR APPLICATION NUMBER: 08/036,210  
; PRIOR FILING DATE: 1995-03-23  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 21  
; LENGTH: 3973  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PTPS31-RD#2  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (1)..(3973)  
; OTHER INFORMATION: n = unknown nucleotide  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2694)

US-10-314-232-21

Query Match 36.8%; Score 2618.6; DB 15; Length 3973;  
Best Local Similarity 97.2%; Pred. No. 0;  
Matches 2716; Conservative 0; Mismatches 4; Indels 75; Gaps 2;

QY 4386 AACAGTCCAGTGTCCACAAATATGCTTTTCTGATGTTCAAGTCAACTAGTGCAC 4445  
DB 33 AAAAGTTCACAGTGTCCACAAATATGCTTTTCTGATGTTCAAGTCAACTAGTGCAC 92  
QY 4446 ATTGACATGATGAAGACCTGACACTATCTCTGGCTACTTTTCAAAATTTACAAATTTACCA 4505  
DB 93 ATTGACATGATGAAGACCTGACACTATCTCTGGCTACTTTTCAAAATTTACAAATTTACCA 152  
QY 4506 TCAACTCTGCTCAAAAATCAAGAATGGAAATCCGAAGATGTGTTGAATATCAAAA 4565  
DB 153 TCAACTCTGCTCAAAAATCAAGAATGGAAATCCGAAGATGTGTTGAATATCAAAA 212  
QY 4566 AATTCATACCTCTATGAAGCTCACTTAATCAAGACAGATATATGATTTAAAGAAATTT 4625  
DB 213 AATTCATACCTCTATGAAGCTCACTTAATCAAGACAGATATATGATTTAAAGAAATTT 272  
QY 4626 TAGATGATATAGATTCGAAGTGGCTGCGACCAATGCTGGCTATGGCAATGCTTCAAA 4685  
DB 273 TAGATGATATAGATTCGAAGTGGCTGCGACCAATGCTGGCTATGGCAATGCTTCAAA 332  
QY 4686 CTGGATTTCTCAAAAACCTCTGCTGGCCCTCCAGATGCTCTCTGAAAATGTTTCATGT 4745  
DB 333 CTGGATTTCTCAAAAACCTCTGCTGGCCCTCCAGATGCTCTCTGAAAATGTTTCATGT 392  
QY 4746 AGTAGCAATCACTTTTAGCATCAGCATAGCTGAGTGAACCTGCTGATTAAGTGG 4805  
DB 393 AGTAGCAATCACTTTTAGCATCAGCATAGCTGAGTGAACCTGCTGATTAAGTGG 452  
QY 4806 ACCACATGTTATCTGATGATGTCAAATCGGTGATGATGATGATGATGATGATGATGATGAT 4865  
DB 453 ACCACATGTTATCTGATGATGTCAAATCGGTGATGATGATGATGATGATGATGATGATGAT 512  
QY 4866 CATCAAGTCAAAATGAAGAAATTAACCATAGAAATTAAGAAATTAAGAAATTAAGAAATTA 4925  
DB 513 CATCAAGTCAAAATGAAGAAATTAACCATAGAAATTAAGAAATTAAGAAATTAAGAAATTA 572  
QY 4926 GTATTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4985  
DB 573 GTATTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 632  
QY 4986 GTCAAGTGTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5045  
DB 633 GTCAAGTGTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 692  
QY 5046 CATGACATTTCAAGAGATACAGATGAAAGTTACAAATTTCAAAATTTCAAAATTTCAAAAT 5105  
DB 693 CATGACATTTCAAGAGATACAGATGAAAGTTACAAATTTCAAAATTTCAAAATTTCAAAAT 752  
QY 5106 TTCTCAACCTAATGGAATATCAAGTATATCAAGTCTGTTTACCGAGAGATGATCC 5165  
DB 753 TTCTCAACCTAATGGAATATCAAGTATATCAAGTCTGTTTACCGAGAGATGATCC 812  
QY 5166 TACTGCTGTGAGATTTCAACCTCAGATTTATACAGAAACCAACATTCGTCATTCG 5225  
DB 813 TACTGCTGTGAGATTTCAACCTCAGATTTATACAGAAACCAACATTCGTCATTCG 872  
QY 5226 AATGCTAGAGACTTAAAGGTGGACATACATACATATCAGTGTTTAGCGAGTCAATAG 5285  
DB 873 AATGCTAGAGACTTAAAGGTGGACATACATACATATCAGTGTTTAGCGAGTCAATAG 932  
QY 5286 TGCTGGTGCAGGTCCAAAGTTCGATGAGAAATACCAAGTATCAAAAGTCTCCAGCAG 5345  
DB 933 TGCTGGTGCAGGTCCAAAGTTCGATGAGAAATACCAAGTATCAAAAGTCTCCAGCAG 992  
QY 5346 ACCAAAAACCAACCAACCCCTATTATGATGCCACAGAAATCTGCTTGTGACTTCAAC 5405  
DB 993 ACCAAAAACCAACCAACCCCTATTATGATGCCACAGAAATCTGCTTGTGACTTCAAC 1052

QY 5406 AACAAATTAATCAGAATGCCAAATATGTTACTACAGTGAATGATGATGATGATGATGATGATGAT 5465  
DB 1053 AACAAATTAATCAGAATGCCAAATATGTTACTACAGTGAATGATGATGATGATGATGATGATGAT 1112  
QY 5466 TGTACAAGTCTTGGCAGCAGAAACAGGAGCTCAGCATGATGATGATGATGATGATGATGATGATGAT 5525  
DB 1113 TGTACAAGTCTTGGCAGCAGAAACAGGAGCTCAGCATGATGATGATGATGATGATGATGATGATGAT 1172  
QY 5526 TGATGCATATTTTAAATAAGCAAGGCCATATTTTAAATAAGCAAGGCCCTTTCTTAACCCCTCC 5585  
DB 1173 TGATGCATATTTTAAATAAGCAAGGCCATATTTTAAATAAGCAAGGCCCTTTCTTAACCCCTCC 1232  
QY 5586 ATGTAACAGAAAGAAAGACAAAGTTTGTGGCAATGAAGAAATCTACATCATATGATGATGATGATGAT 5645  
DB 1233 ATGTAACAGAAAGAAAGACAAAGTTTGTGGCAATGAAGAAATCTACATCATATGATGATGATGATGAT 1292  
QY 5646 TAAATGCATGATGATTTCTGGCAATGAAGCAAAATTTGCAATGACCACTGCAACCAAA 5705  
DB 1293 TAAATGCATGATGATTTCTGGCAATGAAGCAAAATTTGCAATGACCACTGCAACCAAA 1352  
QY 5706 AAGCAATATCTTATTTAAATTTAGAGCTACAAATATTTAGGCAAAATTTTACTGACTCTCA 5765  
DB 1353 AAGCAATATCTTATTTAAATTTAGAGCTACAAATATTTAGGCAAAATTTTACTGACTCTCA 1412  
QY 5766 TTATTTCACTGCTGTTAAGACTTTTAGGGAAGGACTTTTCAAGAAAGAAACCGTAGAGATCAT 5825  
DB 1413 TTATTTCACTGCTGTTAAGACTTTTAGGGAAGGACTTTTCAAGAAAGAAACCGTAGAGATCAT 1472  
QY 5826 TCTTTCCGTCACCTTTGCTGATCTCTTCAATTAATCTCTTGGAAACAGCTATTTTGTGCAAT 5885  
DB 1473 TCTTTCCGTCACCTTTGCTGATCTCTTCAATTAATCTCTTGGAAACAGCTATTTTGTGCAAT 1532  
QY 5886 TGCAGAAATTTGACAGAGAGCAAGAAAGAGTGGCACTACTCTCTCAGAGATGCAAGAAAT 5945  
DB 1533 TGCAGAAATTTGACAGAGAGCAAGAAAGAGTGGCACTACTCTCTCAGAGATGCAAGAAAT 1592  
QY 5946 TATTGACACTAAATTTGAAGCTGGATCAGCTCATCTACAGTGGCAGACCTTGGAACTGAAGGA 6005  
DB 1593 TATTGACACTAAATTTGAAGCTGGATCAGCTCATCTACAGTGGCAGACCTTGGAACTGAAGGA 1652  
QY 6006 CGAGAGATTAAGCGG----- 6020  
DB 1653 CGAGAGATTAAGCGGATACTCTTCAATTTTCTTTAGACGCAAGGAGATTTTGTGATCCCA 1712  
QY 6021 -----GCCAATAAGCAAGAAATCCTTCTCTCCTCAACATCT 6053  
DB 1713 GTTACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1772  
QY 6054 TGAAGAGCTTTGCAACAAACCAACCTTAAAGTTTCAAGAAAGAAATTTTGGAAATTTACCAAA 6113  
DB 1773 TGAAGAGCTTTGCAACAAACCAACCTTAAAGTTTCAAGAAAGAAATTTTGGAAATTTACCAAA 1832  
QY 6114 ATTTCTTCAGGATCTTTCTTCAACTGATGCTGATCTGCTTGGAAATAGAGCAAAACCG 6173  
DB 1833 ATTTCTTCAGGATCTTTCTTCAACTGATGCTGATCTGCTTGGAAATAGAGCAAAACCG 1892  
QY 6174 TTTTCCCAACATAAACCATATATAATAATAACAGAGTAAAGCTGATGATGATGATGATGATGATGAT 6233  
DB 1893 CTTTCCCAACATAAACCATATATAATAATAACAGAGTAAAGCTGATGATGATGATGATGATGATGAT 1949  
QY 6234 TGTTCAGGTTCCGATTTATTTAATGCGCAGCTATATTTCTGTTTATTTATGTTCCAAATGCA 6293  
DB 1950 TGTTCAGGTTCCGATTTATTTAATGCGCAGCTATATTTCTGTTTATTTATGTTCCAAATGCA 2009  
QY 6294 ATTTATTTGCTACTCAAGTCCACTACAGGAAACAGTTGAGAGATTTTGGAGAAATGCTGTG 6353  
DB 2010 ATTTATTTGCTACTCAAGTCCACTACAGGAAACAGTTGAGAGATTTTGGAGAAATGCTGTG 2069  
QY 6354 GGAACACAGGCAAAACATTTAGTAACTGATCAACAGTGTGTTTGAAGAGGACGATCAG 6413  
DB 2070 GGAACACAGGCAAAACATTTAGTAACTGATCAACAGTGTGTTTGAAGAGGACGATCAG 2129

QY 6414 ATGCCATCATGTTGGCCAGAGGACAAACAGCCAGTTACTGCTTTGGAGATATAGTAT 6473  
Db ATGCCATCATGTTGGCCAGAGGACAAACAGCCAGTTACTGCTTTGGAGATATAGTAT 2189  
QY 6474 TACAAAGCTAATGGAGGATGTTCAATAGATTGGACTATCAGGGATCTGAAAAATTGAAAG 6533  
Db TACAAAGCTAATGGAGGATGTTCAATAGATTGGACTATCAGGGATCTGAAAAATTGAAAG 2249  
QY 6534 GCATGGGATTCATGACTGTTCCGACAGTGAACCTTTACTGCTGGCCAGAGCATGGGT 6593  
Db GCATGGGATTCATGACTGTTCCGACAGTGAACCTTTACTGCTGGCCAGAGCATGGGT 2309  
QY 6594 TCTCGAAGACAGGCGCCCTCTAATACCTTTCTGGAAGTTGGTTCCAGCAAGCAGGCGACA 6653  
Db TCTCGAAGACAGGCGCCCTCTAATACCTTTCTGGAAGTTGGTTCCAGCAAGCAGGCGACA 2369  
QY 6654 TGACACCACTATGATTGTTCTACGACGCTGCTGGAGTTGGAAGAACTGGAGTTTAT 6713  
Db TGACACCACTATGATTGTTCTACGACGCTGCTGGAGTTGGAAGAACTGGAGTTTAT 2429  
QY 6714 TGCTCTGGACCAATTAACACACATATTAATGACCATGATTTTGGATATATATGGACT 6773  
Db TGCTCTGGACCAATTAACACACATATTAATGACCATGATTTTGGATATATATGGACT 2489  
QY 6774 AGTAGCTGAACGAGAGTGAAGAAATGTGCATGTGCGAGAACTCGGCACAGTATATCTT 6833  
Db AGTAGCTGAACGAGAGTGAAGAAATGTGCATGTGCGAGAACTCGGCACAGTATATCTT 2549  
QY 6834 TTTCACCACTGCTATGCTGATCTCTTATCAATTAAGGAAAGTAACTAGCCCATCTGTTT 6893  
Db TTTCACCACTGCTATGCTGATCTCTTATCAATTAAGGAAAGTAACTAGCCCATCTGTTT 2609  
QY 6894 TGTTAACTATTCAGCACTTCAGAACTGCTTTTGGACGCTATGGAAGGATGTTGA 6953  
Db TGTTAACTATTCAGCACTTCAGAACTGCTTTTGGACGCTATGGAAGGATGTTGA 2669  
QY 6954 GCTTGAATGGGAAGAAACACATATCTAAATATTTCAGACCAAGGATACAAATGGAAGAGA 7013  
Db GCTTGAATGGGAAGAAACACATATCTAAATATTTCAGACCAAGGATACAAATGGAAGAGA 2729  
QY 7014 TTTTAAATCCAGGGGCAAAAGTTACCCCTCATCTTCCGAATGTAATGTGCAACT 7073  
Db TTTTAAATCCAGGGGCAAAAGTTACCCCTCATCTTCCGAATGTAATGTGCAACT 2789  
QY 7074 TAAAGAAATATCTATGCTTCTCTCACTGCGCTTT 7108  
Db TAAAGAAATATCTATGCTTCTCTCACTGCGCTTT 2824

## RESULT 9

US-10-087-684-7

; Sequence 7, Application US/10087684

; Publication No. US20040029116A1

; GENERAL INFORMATION:

; APPLICANT: Edinger, Shlomit R.  
; APPLICANT: MacDougall, John R.  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Stone, David J.  
; APPLICANT: Grosse, William M.  
; APPLICANT: Lepley, Denise M.  
; APPLICANT: Rieger, Daniel K.  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Casman, Stacie, J.  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Boldog, Ferenc L.  
; APPLICANT: Li, Li  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Mishra, Vishnu  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Tchernov, Velizar T.  
; APPLICANT: Vernet, Corine A.M.

; APPLICANT: Zerhusen, Bryan D.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Miller, Charles E.  
; APPLICANT: Gangolli, Esha A.  
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
; FILE REFERENCE: 21402-214 CIP  
; CURRENT APPLICATION NUMBER: US/10/087,684  
; CURRENT FILING DATE: 2003-03-10  
; PRIOR APPLICATION NUMBER: 60/253,834  
; PRIOR FILING DATE: 2000-11-29  
; PRIOR APPLICATION NUMBER: 60/250,926  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: 60/264,180  
; PRIOR FILING DATE: 2001-01-25  
; PRIOR APPLICATION NUMBER: 60/274,194  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/313,656  
; PRIOR FILING DATE: 2001-08-20  
; PRIOR APPLICATION NUMBER: 60/327,456  
; PRIOR FILING DATE: 2001-10-05  
; NUMBER OF SEQ ID NOS: 220  
; SOFTWARE: CuraSeqList version 0.1  
; SEQ ID NO 7  
; LENGTH: 2565  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2565)  
US-10-087-684-7

Query Match 35.9%; Score 2551; DB 17; Length 2565;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2554; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3262 CCTGAAGGTTGTTGGAAACCTGACTTACCAATCCATTTCTGCTCAATGCAATAAATGTA 3321  
Db CCTGAAGGTTGTTGGAAACCTGACTTACCAATCCATTTCTGCTCAATGCAATAAATGTA 66  
QY 3322 AGCTGGGTCCCAACCGGCTCAACCAACCGGTCTAGTCTTCTACTATGTTTCACTGATCTTA 3381  
Db AGCTGGGTCCCAACCGGCTCAACCAACCGGTCTAGTCTTCTACTATGTTTCACTGATCTTA 126  
QY 3382 CAGCAGACTCTCCGCTGAGACCACTCTTGTGTACATATGAGAGAGCATATATTT 3441  
Db CAGCAGACTCTCCGCTGAGACCACTCTTGTGTACATATGAGAGAGCATATATTT 186  
QY 3442 GATATCTGGAAATAACACTGATATATATATATAAATAATCTCCATCAACAGAAAGGA 3501  
Db GATATCTGGAAATAACACTGATATATATATAAATAATCTCCATCAACAGAAAGGA 246  
QY 3502 TTCTGTGATCTATATCTGCCAGCTATACATCAAGACTGAAGAGATGTCCAGAAACT 3561  
Db TTCTGTGATCTATATCTGCCAGCTATACATCAAGACTGAAGAGATGTCCAGAAACT 306  
QY 3562 TCACCAATAACACACTTTTAAACCTTTTCCCTACTCAGTCTCTTATCATGGAT 3621  
Db TCACCAATAACACACTTTTAAACCTTTTCCCTACTCAGTCTCTTATCATGGAT 366  
QY 3622 CCCCCAGTAAAGCCAAATGGTGCAATAAATAGTTATGATTTAACTTTTAAAGGACCAAT 3681  
Db CCCCCAGTAAAGCCAAATGGTGCAATAAATAGTTATGATTTAACTTTTAAAGGACCAAT 426  
QY 3682 GAAAAATTATCTTTCAATTAATCTTGTGATTAATTAATAATTTGGAAGAGCTTTCCACTTT 3741  
Db GAAAAATTATCTTTCAATTAATCTTGTGATTAATTAATAATTTGGAAGAGCTTTCCACTTT 486  
QY 3742 ACATATATAGCTTTTGTGCTCCGCAAGACTAGAAAGCACTTGGTCTTCCAGTAT 3801  
Db ACATATATAGCTTTTGTGCTCCGCAAGACTAGAAAGCACTTGGTCTTCCAGTAT 546  
QY 3802 CTTTCTTTTACACAGATGAGTCAGTCCGCTTAGCACCTCCACAAAATTTGACTTTAATC 3861

|||||  
547 CTTTTCTTTTACACAGTGGAGTGGCGCTGTAGCACCTCCACAAAATTTGACTTTAATC 606  
Db  
3862 AACTGTACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTCCAGGTGTATT 3921  
Qy  
607 AACTGTACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTCCAGGTGTATT 666  
Db  
3922 GTTAAAGTATATAGTTTAAATTCATGAACATGAACATGACATATATATTATAGAAAT 3981  
Qy  
667 GTTAAAGTATATAGTTTAAATTCATGAACATGAACATGACATATATATTATAGAAAT 726  
Db  
3982 ATATCAGGATTTAAACATGAAGCAAACTTTGTGGACTGGAACCCAGTCAGCACTACTCT 4041  
Qy  
727 ATATCAGGATTTAAACATGAAGCAAACTTTGTGGACTGGAACCCAGTCAGCACTACTCT 786  
Db  
4042 ATCCCGTGTATCTGCGTTTCAACAAAGTTGGAATGGAATCAATTTAGTAAATAGTAA 4101  
Qy  
787 ATCCCGTGTATCTGCGTTTCAACAAAGTTGGAATGGAATCAATTTAGTAAATAGTAA 846  
Db  
4102 TTCAACAAACCAAGATCAGTTCCAGATGTCGTCAGAAATATGCAATGCGAACTAGC 4161  
Qy  
847 TTCAACAAACCAAGATCAGTTCCAGATGTCGTCAGAAATATGCAATGCGAACTAGC 906  
Db  
4162 TGGCAGTCAGTTTATAGTGAATGGGATCCACCBAAGGCAAAATGGAATATAGCGAG 4221  
Qy  
907 TGGCAGTCAGTTTATAGTGAATGGGATCCACCBAAGGCAAAATGGAATATAGCGAG 966  
Db  
4222 TATATGGTAAACAGTTGAAAGGAATTTCAAAAAGTTTCTCCCAAGATCACAATGTACACT 4281  
Qy  
967 TATATGGTAAACAGTTGAAAGGAATTTCAAAAAGTTTCTCCCAAGATCACAATGTACACT 1026  
Db  
4282 TTCAAAAAGCTTTCTGGCAATACCTCATATGTCTTTAAAGTAAAGCTTTCAACCTCAGCT 4341  
Qy  
1027 TTCAAAAAGCTTTCTGGCAATACCTCATATGTCTTTAAAGTAAAGCTTTCAACCTCAGCT 1086  
Db  
4342 GGTGAAGGTGATGAAGCAATGCGCATGTGAGCACTACCTGAAACAGTTCCCGAGTGT 4401  
Qy  
1087 GGTGAAGGTGATGAAGCAATGCGCATGTGAGCACTACCTGAAACAGTTCCCGAGTGT 1146  
Db  
4402 CCCACAAATATGCTTTTCTGATGTTCACTGATGTTCACTGATGCAACATTTAGTAAAG 4461  
Qy  
1147 CCCACAAATATGCTTTTCTGATGTTCACTGATGTTCACTGATGCAACATTTAGTAAAG 1206  
Db  
4462 CCTGACACTATCTTGGCTACTTTTCAAAAATTAACAAATTTACCACTCAACTTGGTCTCAA 4521  
Qy  
1207 CCTGACACTATCTTGGCTACTTTTCAAAAATTAACAAATTTACCACTCAACTTGGTCTCAA 1266  
Db  
4522 AAATGCAAGAAATGGGAATCCGGAAGATGTTGTTGAATATCAAAAATTTCAATACCTCTAT 4581  
Qy  
1267 AAATGCAAGAAATGGGAATCCGGAAGATGTTGTTGAATATCAAAAATTTCAATACCTCTAT 1326  
Db  
4582 GAAGCTCACTTAAGTGAAGCACTGATATGATTAAGAAATTTAGATGATATAGATTC 4641  
Qy  
1327 GAAGCTCACTTAAGTGAAGCACTGATATGATTAAGAAATTTAGATGATATAGATTC 1386  
Db  
4642 CAAGTGGCTGCCAGCACTGCTGCTATGCAATGCTTTCAAACTGGATTTCTACAAA 4701  
Qy  
1387 CAAGTGGCTGCCAGCACTGCTGCTATGCAATGCTTTCAAACTGGATTTCTACAAA 1446  
Db  
4702 ACTCTGCTGCTGCCCTCCAGATGCTCTCTGAAAATGTTTCAATGATAGCAACATCACT 4761  
Qy  
1447 ACTCTGCTGCTGCCCTCCAGATGCTCTCTGAAAATGTTTCAATGATAGCAACATCACT 1506  
Db  
4762 TTTAGCATCAGATAGCTGGAGTGAACCTGCTGCTCAATTTAGTGGACCAATGTTATCTG 4821  
Qy  
1507 TTTAGCATCAGATAGCTGGAGTGAACCTGCTGCTCAATTTAGTGGACCAATGTTATCTG 1566  
Db  
4822 ATTTGATGCAATCGTATGATATGATGATTTTAAATATCTCTCATCAAGTCAATGAA 4881  
Qy  
1567 ATTTGATGCAATCGTATGATATGATGATTTTAAATATCTCTCATCAAGTCAATGAA 1626  
Db  
4882 GAAAAATAAACCATAGAAATTAAGATTTTAAAGATTTTAAAGATTTTCTGTAGTGATC 4941  
Qy

1627 GAAATAAAACCATAGAAATTTAAAGATTTAGAAATTTTCAAGGTATTTCTGTAGTGATC 1686  
Db  
4942 ACTGCATTTACTGGAAACATTTAGTGTCTGCATATGTAGAGGGAAGTCAAGTGTGAAATG 5001  
Qy  
1687 ACTGCATTTACTGGAAACATTTAGTGTCTGCATATGTAGAGGGAAGTCAAGTGTGAAATG 1746  
Db  
5002 ATTTGTTACTTCTTTAGAAATCAGCCCAAGGAGCCCACTTAAACAATGACATTTCCAGAG 5061  
Qy  
1747 ATTTGTTACTTCTTTAGAAATCAGCCCAAGGAGCCCACTTAAACAATGACATTTCCAGAG 1806  
Db  
5062 ATACCAAGATGAAGTTTACAAAATTTTCAATTTAAAGTCTTCTCTCTCTCAACTAATGGA 5121  
Qy  
1807 ATACCAAGATGAAGTTTACAAAATTTTCAATTTAAAGTCTTCTCTCTCTCAACTAATGGA 1866  
Db  
5122 AATATCCAAAGTATATCAAGCTCTGTTTACCGAAGAGATGATCTCTCTCTCTCTCAAGATT 5181  
Qy  
1867 AATATCCAAAGTATATCAAGCTCTGTTTACCGAAGAGATGATCTCTCTCTCTCTCAAGATT 1926  
Db  
5182 CACAACCTCAGTTTATACAGNAACCAACACATTTGCTCATTTGCAATGCTAGAGGACTA 5241  
Qy  
1927 CACAACCTCAGTTTATACAGNAACCAACACATTTGCTCATTTGCAATGCTAGAGGACTA 1986  
Db  
5242 AAGGTGTCAGATATACATATCAATATCAGTGTGTTTACGCAAGTCAATAGTGTCTGTCAGGTCCA 5301  
Qy  
1987 AAGGTGTCAGATATACATATCAATATCAGTGTGTTTACGCAAGTCAATAGTGTCTGTCAGGTCCA 2046  
Db  
5302 AAGGTTCGATGAGAAATACCAATGATATCAAAAGTCCAGCAACCAAAAACCAAAACCA 5361  
Qy  
2047 AAGGTTCGATGAGAAATACCAATGATATCAAAAGTCCAGCAACCAAAAACCAAAACCA 2106  
Db  
5362 ACCCTCTATTTATGATGCCACAGGAAAACCTGTTGTGATCTTCAACAACAAATTTCAATCAGA 5421  
Qy  
2107 ACCCTCTATTTATGATGCCACAGGAAAACCTGTTGTGATCTTCAACAACAAATTTCAATCAGA 2166  
Db  
5422 ATGCCAAATATGTTTACTGATGATGATCATGCAACCAATTAATAAATGTAAGTGTCTGCG 5481  
Qy  
2167 ATGCCAAATATGTTTACTGATGATGATCATGCAACCAATTAATAAATGTAAGTGTCTGCG 2226  
Db  
5482 ACAGAAACAGGAGCTCAGCATGATGGAATTTGAAATGTAACAAAGTGTGATGATGATTTTAAT 5541  
Qy  
2227 ACAGAAACAGGAGCTCAGCATGATGGAATTTGAAATGTAACAAAGTGTGATGATGATTTTAAT 2286  
Db  
5542 AAGCAAGGCCATATTTTCAAAATGAAGCTTTTCTTAAACCTCCATGTAAGAGGAAAG 5601  
Qy  
2287 AAGCAAGGCCATATTTTCAAAATGAAGCTTTTCTTAAACCTCCATGTAAGAGGAAAG 2346  
Db  
5602 ACAGAAATTTAGTGGCAATGNAAGAAATCTACATCATAGTGTGATGATGATGATGATGAT 5661  
Qy  
2347 ACAGAAATTTAGTGGCAATGNAAGAAATCTACATCATAGTGTGATGATGATGATGATGAT 2406  
Db  
5662 CTTGGCAATGAAGACAAAATTTGCAATGGACCACTGAAACCAAAAAGCAATATCTTATTT 5721  
Qy  
2407 CTTGGCAATGAAGACAAAATTTGCAATGGACCACTGAAACCAAAAAGCAATATCTTATTT 2466  
Db  
5722 AAATTTAGAGCTTACAAATATTTATGGGCAATTTTACTGACTCTGATTTATTTCTGACCTGTT 5781  
Qy  
2467 AAATTTAGAGCTTACAAATATTTATGGGCAATTTTACTGACTCTGATTTATTTCTGACCTGTT 2526  
Db  
5782 AAGACTTTAGGGGAAGGACTTTTCAAGAAAGAACCTGTAGAG 5820  
Qy  
2527 AAGACTTTAGGGGAAGGACTTTTCAAGAAAGAACCTGTAGAG 2565  
Db

## RESULT 10

US-10-218-779-7

; Sequence 7, Application US/10218779

; Publication No. US2004002922A1

; GENERAL INFORMATION:

; APPLICANT: Edinger, Shlomit

; APPLICANT: MacDougall, John

; APPLICANT: Millet, Isabelle

; APPLICANT: Ellerman, Karen

; APPLICANT: Stone, David





Db 1447 ACTCTGCTGGCCCTCCAGATGCTCTCTGAAATGTTTCATGTAGTACCAATCACCT 1506  
Qy 4762 TTTAGCATCAGCATAGCTGGAGTGAACTCTGCTGTCATTTACTGAGCAACATGTTATCTG 4821  
Db 1507 TTTAGCATCAGCATAGCTGGAGTGAACTCTGCTGTCATTTACTGAGCAACATGTTATCTG 1566  
Qy 4822 ATTGATGTCAAATCGGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4881  
Db 1567 ATTGATGTCAAATCGGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1626  
Qy 4882 GAAATATAAACCATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4941  
Db 1627 GAAATATAAACCATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1686  
Qy 4942 ACTGCATTTACTGGGAACATTTAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 5001  
Db 1687 ACTGCATTTACTGGGAACATTTAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1746  
Qy 5002 ATTGTTACTACTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5061  
Db 1747 ATTGTTACTACTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1806  
Qy 5062 ATACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5121  
Db 1807 ATACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1866  
Qy 5122 AATATCAAGTATATCAAGCTCTGTTTACCGAGAAGATGATGATGATGATGATGATGATGATGAT 5181  
Db 1867 AATATCAAGTATATCAAGCTCTGTTTACCGAGAAGATGATGATGATGATGATGATGATGATGAT 1926  
Qy 5182 CACAACCTCAGTATATACAGAAACCAACATGCTGTCATTTGCAATGCTAGAGACTA 5241  
Db 1927 CACAACCTCAGTATATACAGAAACCAACATGCTGTCATTTGCAATGCTAGAGACTA 1986  
Qy 5242 AAGGTGGACATACATACATATCAGTGTGTTAGCAGTCAATAGTGTGCTGTCAGTCCA 5301  
Db 1987 AAGGTGGACATACATACATATCAGTGTGTTAGCAGTCAATAGTGTGCTGTCAGTCCA 2046  
Qy 5302 AAGGTTCGGATGAGATACCAATGATATCAAGCTTCCAGCAGCAGCAAAACCAACCA 5361  
Db 2047 AAGGTTCGGATGAGATACCAATGATATCAAGCTTCCAGCAGCAGCAAAACCAACCA 2106  
Qy 5362 ACCCTATTTATGATGCCACAGAAACCTGTTTGTGATCTTCAACCAATTAATCAATCAGA 5421  
Db 2107 ACCCTATTTATGATGCCACAGAAACCTGTTTGTGATCTTCAACCAATTAATCAATCAGA 2166  
Qy 5422 ATGCCCAATGTTACTACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 5481  
Db 2167 ATGCCCAATGTTACTACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2226  
Qy 5482 ACAGAAACAGGAGCTCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5541  
Db 2227 ACAGAAACAGGAGCTCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2286  
Qy 5542 AAGCAAGGCCATATTTTCAATGAGGCTTTCTTAACCTCCATGTCAGAGGAAAG 5601  
Db 2287 AAGCAAGGCCATATTTTCAATGAGGCTTTCTTAACCTCCATGTCAGAGGAAAG 2346  
Qy 5602 ACAGAGTTTGTGCAATGAGAAATCTACATGATGATGATGATGATGATGATGATGATGATGAT 5661  
Db 2347 ACAGAGTTTGTGCAATGAGAAATCTACATGATGATGATGATGATGATGATGATGATGATGAT 2406  
Qy 5662 CTGGCAATGAAGCAAAATTTGCAATGAGCACTGAAACCAAAAGCAATACATTTATTT 5721  
Db 2407 CTGGCAATGAAGCAAAATTTGCAATGAGCACTGAAACCAAAAGCAATACATTTATTT 2466  
Qy 5722 AAATTTAGAGCTACAAATATTTAGGACAATTTACTGATCTGATTTATTTCTGACCTGTT 5781  
Db 2467 AAATTTAGAGCTACAAATATTTAGGACAATTTACTGATCTGATTTATTTCTGACCTGTT 2526  
Qy 5782 AAGACTTTAGGGGAGGACTTTTCAGAAAGACCGTAGAG 5820  
Db 2527 AAGACTTTAGGGGAGGACTTTTCAGAAAGACCGTAGAG 2565

## RESULT 11

US-10-314-232-14  
; Sequence 14, Application US/10314232  
; Publication No. US20030138932A1  
; GENERAL INFORMATION:  
; APPLICANT: MOLLER, NIELS P.H.  
; APPLICANT: MOLLER, KARIN B.  
; APPLICANT: ULLRICH, AXEL  
; FILE REFERENCE: 038602/0686  
; CURRENT APPLICATION NUMBER: US/10/314,232  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/361,096  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: 08/449,609  
; PRIOR FILING DATE: 1995-05-24  
; PRIOR APPLICATION NUMBER: 08/036,210  
; PRIOR FILING DATE: 1995-03-23  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 14  
; LENGTH: 2692  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (1)..(2692)  
; OTHER INFORMATION: n = unknown nucleotide  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (92)..139, 259..1414)  
US-10-314-232-14

Query Match 15.7%; Score 1118.8; DB 15; Length 2692;

Best Local Similarity 93.7%; Pred. No. 7.6e-256;

Matches 1218; Conservative 0; Mismatches 7; Indels 75; Gaps 2;

Qy 5881 GCATTTGCAAGAAATTCGACAGAGCAGAAAGAGTGGCACATATCTCTCAGGATGCA 5940  
Db 247 GCTGTTTAGAATTCGACAGAGCAGAAAGAGTGGCACATATCTCTCAGGATGCA 306  
Qy 5941 GAAATTTATGACACTAAATTTGAAGCTGGATCAGCTCATCACAGTGGCAGACCTGGAATG 6000  
Db 307 GAAATTTATGACACTAAATTTGAAGCTGGATCAGCTCATCACAGTGGCAGACCTGGAATG 366  
Qy 6001 AAGGACGAGAGATTAACGGG----- 6020  
Db 367 AAGGACGAGAGATTAACGGGATCTCTTCATTTTCTTTAGACGCAAGAGATTTTGTGTC 426  
Qy 6021 -----GCCAATAAGCAAGAAATCTCTTCCTGCAA 6048  
Db 427 ATCCAGTTACTTAGTTATAGAAATCCATCAAGCCCAATAGCAAGAAATCTCTTCGCAA 486  
Qy 6049 CATGTTGAAGAGCTTTGCAACAAACCAACCTAAAGTTTCAAGAAAGAAATTTTCGGAATTA 6108  
Db 487 CATGTTGAAGAGCTTTGCAACAAACCAACCTAAAGTTTCAAGAAAGAAATTTTCGGAATTA 546  
Qy 6109 CCAAAATTTCTTCAGGATCTTTCTTCAACTGATGCTGATCTGCTTGGAAATAGAGCAAAA 6168  
Db 547 CCAAAATTTCTTCAGGATCTTTCTTCAACTGATGCTGATCTGCTTGGAAATAGAGCAAAA 606  
Qy 6169 AACGTTTCCCAACATAAACCATATATATATACAGAGTAAAGCTGATAGCTGAC 6228  
Db 607 AACGTTTCCCAACATAAACCATATATATATACAGAGTAAAGCTGATAGCTGAC 663  
Qy 6229 GCTAGTGTTCAGAGTTTCGGATTTATATTAATGCGCAGCTATATTTCTGTTTATTTATGTC 6288  
Db 664 GCTAGTGTTCAGAGTTTCGGATTTATATTAATGCGCAGCTATATTTCTGTTTATTTATGTC 723  
Qy 6289 AATGAATTTATGCTACTCAAGGTCCTACCTACAGAAACAGTTGGAGATTTTGGGAATG 6348

Db 724 AATGAATTTATTGCTACTCAAGGTCCACTACCAGGAACAGTTGGAGATTTTGGAGAATG 783  
QY 6349 GTGTGGAAACCAAGGCAAAACATTTAGTAATGCTTAACACAGTGTTTGAAAAGACGG 6408  
Db 784 GTGTGGAAACCAAGGCAAAACATTTAGTAATGCTTAACACAGTGTTTGAAAAGACGG 843  
QY 6409 ATCAGATGCCATCAGTATTGGCCAGAGGCAACAAGCCAGTTACTGTCTTTGGAGATATA 6468  
Db 844 ATCAGATGCCATCAGTATTGGCCAGAGGCAACAAGCCAGTTACTGTCTTTGGAGATATA 903  
QY 6469 GTGATTACAAAGCTAATGAGAGATGTTCAAATAGATGGAATCAGGGATCTGAAAATT 6528  
Db 904 GTGATTACAAAGCTAATGAGAGATGTTCAAATAGATGGAATCAGGGATCTGAAAATT 963  
QY 6529 GAAAGCATGGGATGTCATGACTGTTTCACAGTGTAACTTTACTGCCCTGGCCAGAGCAT 6588  
Db 964 GAAAGCATGGGATGTCATGACTGTTTCACAGTGTAACTTTACTGCCCTGGCCAGAGCAT 1023  
QY 6589 GGGGTTCTTGAGAACAGCGCCCTCTTAATTTCACTTTGTGAAGTTGGTTTCGAGCAAGCAGG 6648  
Db 1024 GGGGTTCTTGAGAACAGCGCCCTCTTAATTTCACTTTGTGAAGTTGGTTTCGAGCAAGCAGG 1083  
QY 6649 GCACATGACACCACTATGATGTTCTCACTGAGTGTGGAAGTTGGTTTCGAGCAAGCAGG 6708  
Db 1084 GCACATGACACCACTATGATGTTCTCACTGAGTGTGGAAGTTGGTTTCGAGCAAGCAGG 1143  
QY 6709 TTTATTGCTCTGACCAATTTACACAACTATATAATGACCATGATTTTGTGGATATATAT 6768  
Db 1144 TTTATTGCTCTGACCAATTTACACAACTATATAATGACCATGATTTTGTGGATATATAT 1203  
QY 6769 GGACTAGTAGTGAACCTGAGAAAGTGAAGAAATGTGCATGGTGCAGAACTCTGGCAGCAGTAT 6828  
Db 1204 GGACTAGTAGTGAACCTGAGAAAGTGAAGAAATGTGCATGGTGCAGAACTCTGGCAGCAGTAT 1263  
QY 6829 ATCTTTTACACAGTGCATTTCTGGATCTCTTATCAAAATGAGGAAGTAAATCAGCCCATC 6888  
Db 1264 ATCTTTTACACAGTGCATTTCTGGATCTCTTATCAAAATGAGGAAGTAAATCAGCCCATC 1323  
QY 6889 TGTTTTGTAACTATTCAGCACTTCAGAGATGGACTCTTTTGGACGCCATGGAAGTGAT 6948  
Db 1324 TGTTTTGTAACTATTCAGCACTTCAGAGATGGACTCTTTTGGACGCCATGGAAGTGAT 1383  
QY 6949 GTTGAGCTTGAATGGGAAGAAACCACTATGTAATAATTCAGACCAAGGATCAATTTGGA 7008  
Db 1384 GTTGAGCTTGAATGGGAAGAAACCACTATGTAATAATTCAGACCAAGGATCAATTTGGA 1443  
QY 7009 AGAGATTTTAAATCCAGGGGCCAAAGTTACCCCTCATTTCTTCGGAATGGAATGTGC 7068  
Db 1444 AGAGATTTTAAATCCAGGGGCCAAAGTTACCCCTCATTTCTTCGGAATGGAATGTGC 1503  
QY 7069 AACCTTAAGAAATATCTATGCTTCTCTCACTGTGCCTTT 7108  
Db 1504 AACCTTAAGAAATATCTATGCTTCTCTCACTGTGCCTTT 1543

RESULT 12

US-10-314-232-10  
; Sequence 10, Application US/10314232  
; Publication No. US20030138932A1  
; GENERAL INFORMATION:  
; APPLICANT: MOLLER, NIELS P.H.  
; APPLICANT: MOLLER, KARIN B.  
; APPLICANT: ULLRICH, AXEL  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE  
; FILE REFERENCE: 038602/0686  
; CURRENT APPLICATION NUMBER: US/10/314,232  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/361,096  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: 08/449,609  
; PRIOR FILING DATE: 1995-05-24  
; PRIOR APPLICATION NUMBER: 08/036,210  
; PRIOR FILING DATE: 1995-03-23

; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: PatentIn Ver. 2.1.1  
; SEQ ID NO 10  
; LENGTH: 2309  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; OTHER INFORMATION: Description of Artificial Sequence: PTP-S31D  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (1)..(2309)  
; OTHER INFORMATION: n = unknown nucleotide  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (65)..(1030)  
US-10-314-232-10  
Query Match 15.1%; Score 1071.6; DB 15; Length 2309;  
Best Local Similarity 99.4%; Pred. No. 1.3e-244;  
Matches 1087; Conservative 0; Mismatches 4; Indels 3; Gaps 1;  
QY 6015 AACGGCGCAATAAGCAAGAAATCCTTCTGCAACATGTTGAAGAGCTTTGCACAACAA 6074  
Db 70 AATGAGGCCAATAAGCAAGAAATCCTTCTGCAACATGTTGAAGAGCTTTGCACAACAA 129  
QY 6075 CAACCTAAAGTTTCAAGGAAGAAATTTTCGGAATTACCAAAATTTCTTCAGGATCTTTCTTC 6134  
Db 130 CAACCTAAAGTTTCAAGGAAGAAATTTTCGGAATTACCAAAATTTCTTCAGGATCTTTCTTC 189  
QY 6135 AACTGATGCTGATCTGCCCTTGGAAATAGAGCAAAAAACCGTTTCCAAAACATAAAACCAT 6194  
Db 190 AACTGATGCTGATCTGCCCTTGGAAATAGAGCAAAAAACCGTTTCCAAAACATAAAACCAT - 248  
QY 6195 TAATAATAATAACAGATTAAGCTGATAGCTGACGCTAGTGTTCAGGTTCCGATTTAT 6254  
Db 249 --ATAATAATAACAGATTAAGCTGATAGCTGACGCTAGTGTTCAGGTTCCGATTTAT 306  
QY 6255 TAATCCAGCTATATTTCTGGTTATTTATGTCCTCAATGAATTTATTTGCTACTCAAGTCC 6314  
Db 307 TAATCCAGCTATATTTCTGGTTATTTATGTCCTCAATGAATTTATTTGCTACTCAAGTCC 366  
QY 6315 ACTACCAAGCAAGTGGAGATTTTGGAGATGTTGGGAAACCAAGGCAAAAAACATT 6374  
Db 367 ACTACCAAGCAAGTGGAGATTTTGGAGATGTTGGGAAACCAAGGCAAAAAACATT 426  
QY 6375 AGTAATGCTTAACACAGTGTTTTGAAGAGGCGGATCAGATGCCATCAGTATTTGCCAGA 6434  
Db 427 AGTAATGCTTAACACAGTGTTTTGAAGAGGCGGATCAGATGCCATCAGTATTTGCCAGA 486  
QY 6435 GGCAACCAAGCCAGTTACTGTCTTTGGAGATATAGTGAATTAACAAAGCTAAATGGAGATGT 6494  
Db 487 GGCAACCAAGCCAGTTACTGTCTTTGGAGATATAGTGAATTAACAAAGCTAAATGGAGATGT 546  
QY 6495 TCAAAATAGATTGGACTATCAGGGATCTGAAAATTCGAAAGGCGATGGGATTTGCATGCTGT 6554  
Db 547 TCAAAATAGATTGGACTATCAGGGATCTGAAAATTCGAAAGGCGATGGGATTTGCATGCTGT 606  
QY 6555 TCGACAGTGTAACTTTACTGCTGCGCCAGAGCATGGGTTCTGTGAGACAGGCCCTCT 6614  
Db 607 TCGACAGTGTAACTTTACTGCTGCGCCAGAGCATGGGTTCTGTGAGACAGGCCCTCT 666  
QY 6615 AATTCACTTTTGTGAAGTTGGTTTCGAGCAAGGCGCACATGACACACACACCTATGATTTGT 6674  
Db 667 AATTCACTTTTGTGAAGTTGGTTTCGAGCAAGGCGCACATGACACACACACCTATGATTTGT 726  
QY 6675 TCATGCTGAGTGTGGAGTTGGAAGAACTGGAGTTTTTATTTGCTCTGGACCATTTAACACA 6734  
Db 727 TCATGCTGAGTGTGGAGTTGGAAGAACTGGAGTTTTTATTTGCTCTGGACCATTTAACACA 786  
QY 6735 ACATATAAATGACCAATGATTTTGTGGATATATATGCTAGCTAGCTGAGCTGAGAGTCA 6794  
Db 787 ACATATAAATGACCAATGATTTTGTGGATATATATGCTAGCTAGCTGAGAGTCA 846



```
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/10/314,232
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/361,096
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 532
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PCR fragment
; NAME/KEY: CDS
; LOCATION: (3)..(530)
US-10-314-232-8

Query Match          7.5%; Score 532; DB 15; Length 532;
Best Local Similarity 100.0%; Pred. No. 3.9e-116;
Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6437 ACAACAAGCCAGTACTGCTTTGGAGATATAGTACAAAGCTAATGGAGGATGTTTC 6496
DB 1 ACAACAAGCCAGTACTGCTTTGGAGATATAGTACAAAGCTAATGGAGGATGTTTC 60

QY 6497 AAATAGATTGGACTATCAGGATCTGAAATTTGAAAGGATGGGATTCATGACTGTTTC 6556
DB 61 AAATAGATTGGACTATCAGGATCTGAAATTTGAAAGGATGGGATTCATGACTGTTTC 120

QY 6557 GACAGTGTAACCTTACTGCTGGCCAGCATGGGTTCTGAGAACAGCGCCCTCTAA 6616
DB 121 GACAGTGTAACCTTACTGCTGGCCAGCATGGGTTCTGAGAACAGCGCCCTCTAA 180

QY 6617 TTCACTTTGTGAAGTTGGTTTCGAGCAAGCAGGCGCATGACACCACTATGATTGTTTC 6676
DB 181 TTCACTTTGTGAAGTTGGTTTCGAGCAAGCAGGCGCATGACACCACTATGATTGTTTC 240

QY 6677 ACTCAGTCTCGAGTTGGAAGAACTGGAGTTTTTATTGCTCTGACCACTTTAACACAAC 6736
DB 241 ACTCAGTCTCGAGTTGGAAGAACTGGAGTTTTTATTGCTCTGACCACTTTAACACAAC 300

QY 6737 ATATAAATGACCATGATTTTGTGGATATATATGAGCTAGTACTGACTGAGAGTGAAA 6796
DB 301 ATATAAATGACCATGATTTTGTGGATATATATGAGCTAGTACTGACTGAGAGTGAAA 360

QY 6797 GAATGTGATGTGAGAAATCTGGACAGTATATCTTTTACACAGATGCTATCTTGGATC 6856
DB 361 GAATGTGATGTGAGAAATCTGGACAGTATATCTTTTACACAGATGCTATCTTGGATC 420

QY 6857 TCTTATCAATAAGGGAAGTAATCAGCCCATCTGTTTTGTTAACTATTTCAGCACTTCAGA 6916
DB 421 TCTTATCAATAAGGGAAGTAATCAGCCCATCTGTTTTGTTAACTATTTCAGCACTTCAGA 480

QY 6917 AGATGAGCTCTTTGGACGCCATGGAAGGTGATGTTGAGCTTGAATGGGAGA 6968
DB 481 AGATGAGCTCTTTGGACGCCATGGAAGGTGATGTTGAGCTTGAATGGGAGA 532
```

RESULT 15

US-10-314-232-3

; Sequence 3, Application US/10314232

; Publication NO. US20030138932A1

; GENERAL INFORMATION:

; APPLICANT: MOLLER, NIELS P.H.

```
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/10/314,232
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/361,096
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 321
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PTP-S31
; NAME/KEY: CDS
; LOCATION: (1)..(321)
US-10-314-232-3
```

```
Query Match          4.5%; Score 319.4; DB 15; Length 321;
Best Local Similarity 99.7%; Pred. No. 1.7e-65;
Matches 320; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6355 GAAACCAGGGCAAAAACATTAGTAATGCTAAACACAGTGTGTTTTGAAAAAGGACGGATCAGA 6414
DB 1 GAAACCAGGGCAAAAACATTAGTAATGCTAAACACAGTGTGTTTTGAAAAAGGACGGATCAGA 60

QY 6415 TGCCATCAGTATTGGCCAGAGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGATT 6474
DB 61 TGCCATCAGTATTGGCCAGAGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGATT 120

QY 6475 ACAAAGCTAATGAGGAGATGTTCAAAATAGATTGGACTATCAGGATCTGAAAAATTCGAAAGG 6534
DB 121 ACAAAGCTAATGAGGAGATGTTCAAAATAGATTGGACTATCAGGATCTGAAAAATTCGAAAGG 180

QY 6535 CATGGGGATTGCATGACTGTTTCGACAGTGTAACTTTTACTGCTGGCCAGAGCATGGGGTT 6594
DB 181 CATGGGGATTGCATGACTGTTTCGACAGTGTAACTTTTACTGCTGGCCAGAGCATGGGGTT 240

QY 6595 CCTGAGAACAGCGCCCTCTTAATTCACTTTGTGAAGTTGGTTTCGAGCAAGCAGGCGCAT 6654
DB 241 CCTGAGAACAGCGCCCTCTTAATTCACTTTGTGAAGTTGGTTTCGAGCAAGCAGGCGCAT 300

QY 6655 GACACCACACCTATGATTGTT 6675
DB 301 GACACCACACCTATGATTGTT 321
```

Search completed: July 3, 2005, 10:49:58

Job time : 5493 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2005, 20:48:19 ; Search time 32708 Seconds  
(without alignments)  
8272.008 Million cell updates/sec

Title: US-10-673-885-1  
Perfect score: 7108  
Sequence: 1 taatgtgactgtccagaa.....gcttctcactgtgccttt 7108

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hic: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_gss1: \*  
9: gb\_gss2: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score  | Match | Length | DB | ID       | Description         |
|------------|--------|-------|--------|----|----------|---------------------|
| 1          | 2399.2 | 33.8  | 3416   | 3  | AK041657 | AK041657 Mus muscu  |
| 2          | 625.2  | 8.8   | 761    | 6  | CB167019 | CB167019 BTU602600  |
| 3          | 491    | 6.9   | 561    | 2  | BE668393 | BE668393 157119 MA  |
| 4          | 485.2  | 6.8   | 538    | 2  | BE751055 | BE751055 202768 MA  |
| 5          | 401.8  | 5.7   | 517    | 4  | BM750287 | BM750287 K-EST0025  |
| 6          | 388.4  | 5.5   | 689    | 2  | BB230596 | BB230596 BB230596   |
| 7          | 380    | 5.3   | 605    | 2  | BB638355 | BB638355 BB638355   |
| 8          | 355.4  | 5.0   | 905    | 5  | BU222226 | BU222226 603106463  |
| 9          | 301.6  | 4.2   | 484    | 8  | AQ812654 | AQ812654 HS 5266 B  |
| 10         | 277    | 3.9   | 771    | 9  | CE385339 | CE385339 tigr-gss-  |
| 11         | 230.8  | 3.2   | 412    | 2  | AW081039 | AW081039 xc359f04.x |
| 12         | 215    | 3.0   | 746    | 5  | BU222294 | BU222294 603105362  |
| 13         | 190.8  | 2.7   | 629    | 8  | B66492   | B66492 CIT-HSP-201  |
| 14         | 190.6  | 2.7   | 600    | 3  | CR749546 | CR749546 Homo sapi  |
| 15         | 180.6  | 2.5   | 547    | 8  | BH120957 | BH120957 RPCI-24-2  |
| 16         | 180.6  | 2.5   | 713    | 9  | AG323589 | AG323589 Mus muscu  |
| 17         | 175.8  | 2.5   | 779    | 7  | CF285207 | CF285207 AGENCOURT  |
| 18         | 175    | 2.5   | 756    | 9  | CC914861 | CC914861 E084G17ba  |
| 19         | 169.4  | 2.4   | 921    | 6  | CD300952 | CD300952 AGENCOURT  |
| 20         | 168.2  | 2.4   | 703    | 8  | AQ587839 | AQ587839 CITBI-E1-  |
| 21         | 165.4  | 2.3   | 6038   | 3  | CR749277 | CR749277 Homo sapi  |
| 22         | 161.6  | 2.3   | 853    | 5  | BW359375 | BW359375 BW359375   |
| 23         | 160.8  | 2.3   | 707    | 5  | BW484146 | BW484146 BW484146   |
| 24         | 158    | 2.2   | 583    | 7  | CK315103 | CK315103 SB02041B1  |

|    |       |     |      |   |          |                    |
|----|-------|-----|------|---|----------|--------------------|
| 25 | 155.4 | 2.2 | 858  | 7 | CR428355 | CR428355           |
| 26 | 154.8 | 2.2 | 919  | 5 | BQ710210 | BQ710210 AGENCOURT |
| 27 | 153.8 | 2.2 | 707  | 7 | CN790080 | CN790080 4124633 B |
| 28 | 153.8 | 2.2 | 779  | 7 | CK777687 | CK777687 964818 MA |
| 29 | 153.4 | 2.2 | 744  | 6 | CA512045 | CA512045 UI-R-EJO- |
| 30 | 153.2 | 2.2 | 835  | 6 | CD804595 | CD804595 UI-M-GVO- |
| 31 | 151.8 | 2.1 | 671  | 8 | AQ580682 | AQ580682 RPCI-11-4 |
| 32 | 150.8 | 2.1 | 694  | 6 | CD353637 | CD353637 UI-M-GMO- |
| 33 | 149.8 | 2.1 | 719  | 9 | AG323639 | AG323639 Mus muscu |
| 34 | 148.8 | 2.1 | 1191 | 6 | CD504171 | CD504171 CDA66-E03 |
| 35 | 146.6 | 2.1 | 774  | 4 | BI105833 | BI105833 60289253  |
| 36 | 145.6 | 2.0 | 859  | 7 | CN325188 | CN325188 AGENCOURT |
| 37 | 144.8 | 2.0 | 594  | 7 | CF181137 | CF181137 817938 MA |
| 38 | 144   | 2.0 | 732  | 6 | CD350610 | CD350610 UI-M-GYO- |
| 39 | 144   | 2.0 | 775  | 6 | CA328425 | CA328425 UI-M-FIO- |
| 40 | 143.4 | 2.0 | 664  | 1 | AU206194 | AU206194 AU206194  |
| 41 | 143.2 | 2.0 | 4286 | 9 | AY417567 | AY417567 Homo sapi |
| 42 | 141.8 | 2.0 | 650  | 1 | AU130535 | AU130535 AUL130535 |
| 43 | 141.6 | 2.0 | 701  | 7 | CK807021 | CK807021 AGENCOURT |
| 44 | 141   | 2.0 | 3498 | 3 | AK078884 | AK078884 Mus muscu |
| 45 | 140.8 | 2.0 | 628  | 8 | AQ243425 | AQ243425 HS_2058_A |

## ALIGNMENTS

|            |  |          |         |      |        |                 |
|------------|--|----------|---------|------|--------|-----------------|
| RESULT 1   | AK041657   | AK041657 | 3416 bp | mRNA | linear | HTC 03-APR-2004 |
| LOCUS      | AK041657   |          |         |      |        |                 |
| DEFINITION | Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone: A630028F16 product: hypothetical Fibronectin type III domain/Fibronectin type III repeat containing protein, full insert sequence.   |          |         |      |        |                 |
| ACCESSION  | AK041657   |          |         |      |        |                 |
| VERSION    | AK041657.1 GI:263334640  |          |         |      |        |                 |
| KEYWORDS   | HTC; CAP trapper.  |          |         |      |        |                 |
| SOURCE     | Mus musculus (house mouse)   |          |         |      |        |                 |
| ORGANISM   | Mus musculus   |          |         |      |        |                 |
| REFERENCE  | 1 Carninci, P., and Hayashizaki, Y.  |          |         |      |        |                 |
| AUTHORS    | Carninci, P., and Hayashizaki, Y.  |          |         |      |        |                 |
| TITLE      | High-efficiency full-length cDNA cloning   |          |         |      |        |                 |
| JOURNAL    | Meth. Enzymol. 303, 19-44 (1999)   |          |         |      |        |                 |
| MEDLINE    | 99279253   |          |         |      |        |                 |
| PUBMED     | 10349636   |          |         |      |        |                 |
| REFERENCE  | 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  |          |         |      |        |                 |
| AUTHORS    | Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  |          |         |      |        |                 |
| TITLE      | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes   |          |         |      |        |                 |
| JOURNAL    | Genome Res. 10 (10), 1617-1630 (2000)  |          |         |      |        |                 |
| MEDLINE    | 20499374   |          |         |      |        |                 |
| PUBMED     | 11042159   |          |         |      |        |                 |
| REFERENCE  | 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujitake, S., Inoue, K., Togawa, I., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. |          |         |      |        |                 |
| AUTHORS    | Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujitake, S., Inoue, K., Togawa, I., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.   |          |         |      |        |                 |
| TITLE      | RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer   |          |         |      |        |                 |
| JOURNAL    | Genome Res. 10 (11), 1757-1771 (2000)  |          |         |      |        |                 |
| MEDLINE    | 20530913   |          |         |      |        |                 |
| PUBMED     | 11076861   |          |         |      |        |                 |
| REFERENCE  | 4 The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.   |          |         |      |        |                 |
| AUTHORS    | Functional annotation of a full-length mouse cDNA collection   |          |         |      |        |                 |
| TITLE      | Nature 409, 685-690 (2001)   |          |         |      |        |                 |
| JOURNAL    | Nature 409, 685-690 (2001)   |          |         |      |        |                 |
| REFERENCE  | 5  |          |         |      |        |                 |

## AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

## TITLE

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

## JOURNAL

Nature 420, 563-573 (2002)

## REFERENCE

6 (bases 1 to 3416)

## AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, S., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

## TITLE

Direct Submission

## JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Adenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge. Please visit our web site for further details. URL: <http://genome.gsc.riken.jp/> URL: <http://fantom.gsc.riken.jp/>.

## FEATURES

source

Location/Qualifiers

1..3416

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM DB:A630028F16"

/db\_xref="taxon:10090"

/clone="A630028F16"

/tissue\_type="thymus"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="3 days neonate"

157..>3414

/note="unnamed protein product; hypothetical Fibrinectin type III domain/Fibrinectin type III repeat containing protein (InterPro|IPR003961, InterPro|IPR003962, evidence: putative"

/codon\_start=1

/protein\_id="BAC31021.1"

/db\_xref="GI:26334641"

/translation="MDLFFFLFLIGTSQVDVSGSPDDTVYDITLSSISATTVSS

PVSRILATNVSPGPPVFLAGRVSGAGILLSWNPNNPGRIRIISYVVKYKEVCPMMQ

TAIVRAKPDLSLEVLTLNLPQTVEIKVAENASAGIGVSDPFLFQTAESAPGVV

NLTVALNYSANLILWLPDPNGKITSFKISVGHARSIGIVKVDKSIKVEDLLSGKLP

ENKNSDSFLMSTSPSLSRATPLRTHLSNLTARNKLSISVMKEPISFVYTHLRP

YTYLFEVSAVTEAGIDSTIVRPESVPEGPQNCITGNVIGKAFSLISWDDPALIV

GKFSRVLYGPTGRIDNSTKDLRFTHLPFTFYDYYVAARTSAGVGPKNLSVF

TPDVPVGDQIDQVEATEIRVSRWKPONGIIISOYRVKVSLESGVILENTILT

GODEV LNNPMTWIDMLVDPMIGFVGSGMSDLHSLASFYINSHPHDPFPAITRVR

QSPVYVATNVOYMTDIAEHLISYVIRLVLPETHTISVSFAFTVMGEGPPTLVITRVR

EQVPSIQLINKYINSSSILLYWDPPEYNGKITHYIYAMELTDNRAPQMTVDNS

FLITGLUKYKRYKMRVAASTHVGESSEENDLFRKTPDEPSSPDQVKKVTDVDSFE

LSLWSPPKPKNGIITAYFVYONADALEFVNKSTNITLSDLKPYTLNYSIQSYR

LGHNGQSLLSVRTSETVDPAPENITKYSSEIEIFLPPRSPNGIIIOKTYIL

KRSNSHEARTIETLSLTITGGKIKYTHVIVSASTLKGEGVRSMPI SILTEADAP

## CDS

| ORIGIN |     | Query Match   | 33.8%; Score 2399.2; DB 3; Length 3416;                  |
|--------|-----|---|--|
|        |     | Best Local Similarity   | 84.2%; Pred. No. 0;                                      |
|        |     | Mismatches  | 2754; Conservative 0; Mismatches 508; Indels 10; Gaps 4; |
| QY     | 1   | TTAATTTGTGTACTTGCAGAGGATCTGTCTTTAAATCATTAATGAGGAGCAATTTCTCT   | 60   |
| DB     | 51  | TTAATTTGTGTACTTGCAGAGGATCTGTCTTTAAATCATTAATGAGGAGCAATTTCTCT   | 110  |
| QY     | 61  | CTAGAGCCATCAATGTGATTTCTACTGGCTG-AAAAATGTAATAAGATGGATTTCTTAT   | 119  |
| DB     | 111 | CTAGAGCCATCAATGTGATTTGTACTGTAGAAAAAATGTAATAAGATGGATTTCTTAT    | 170  |
| QY     | 120 | CAATTTTCTTTTACTTTTATTTGGGATTTCTGTCTTTGGGATTTCTGTCTTTGGAAT     | 179  |
| DB     | 171 | CTTTTCTTTTCTCTCTCTCTCATTTGGAATTTCTGTCTTTGGAATTTCTGTCTTTT      | 230  |
| QY     | 180 | TCCTGGTACTAGGTAGGATTAACCATCTCTTCAATTTCT---ACAACATACACCTCACC   | 236  |
| DB     | 231 | TCATGACACCTGTGTATGATATCACGCTGTCTTCAATTTTGTCCCAACACCTACAGTCCAC | 290  |
| QY     | 237 | TCGTTACTAGATAGTGACACCAAAATGTAACAAACACAGGCTCTCAGTCTCTCTACCGG   | 296  |
| DB     | 291 | CGTGTAGTACACCTGTGGCAACAAATGTGTCAAACCGGGCTCTCAGTCTCTCTACCGG    | 350  |
| QY     | 297 | GGAAAGAGTCCGATCTGTCTGGGATTTCTTCTGTCTTTGGAATCACCACTTAATCCAATGG | 356  |
| DB     | 351 | GGAGAGAGTGTGCTCTGCGGGATTTCTCTCTTCATGGAATACCCCTCTTAACCAATGG    | 410  |
| QY     | 357 | AAGGATATATCTTACATTTGTCAAATATAGGAAGTTTGTCCGTGGATGCAACAGTATA    | 416  |
| DB     | 411 | AAGAATCATATCGTATGTTGTTAAATATCAAGGAAGTGTGTCCATGGATGACAGACGTA   | 470  |
| QY     | 417 | TACCAAGTCAGATCAAAAGCCAGACAGTCTGGAAGTTCTTCTTAACTTAATCTCTGG     | 476  |
| DB     | 471 | TACCGAGTGAGAGGAGGAGCCGACAGTCTGGAAGTCTTCTTCACTTAACCTTAAACCTGG  | 530  |
| QY     | 477 | AACAACATATGAAATTAAGTTTGTCTGTAACACAGTCTGCAATTTGGAGTGTGTAGTGA   | 536  |
| DB     | 531 | AACAACATACAGATTAAGTTTGTCTGCGGAAACACGCTGCAATTTGGAGTGTGTAGTGA   | 590  |
| QY     | 537 | TCATTTCTCTTCCAAAATGTCAGAAAGTCTCCAGAAAGTGTGTAATCTCACAGTTGA     | 596  |
| DB     | 591 | CCCATTTCTTTTCAAAACCGCAGAAAGTGTCTCCAGGAAAGTGTGTAATCTCACAGTTGA  | 650  |
| QY     | 597 | GGCTTACACGCTTCCAGCAGTTAAGCTGATTTGTTTACCTCGGCAACCAATGGCAA      | 656  |
| DB     | 651 | GGCTTCAACTATCTGCAAGTAACTTGAATTTGGTATTTTACCTCGGCAACCAATGGTAA   | 710  |
| QY     | 657 | AATACCAGCTTCAAGATTTAGTGTCAAAACATGCCAGAGTGGGATAGTAGTGAAGATGT   | 716  |
| DB     | 711 | AATACCAGCTTCAAGATTTAGTGTCAAAACATGCCAGAGTGGGATAGTAGTGAAGATGT   | 770  |
| QY     | 717 | CTCAATACAGAGTAGGAGCAATTTTGACTGGGAAATTTGCCAGATGCAATGAGATAGTGA  | 776  |
| DB     | 771 | CTCCATCAAGTAGAGGACCTTTTATCTGGGAAATTTACCTGAATGTAATGAGAACAGTGA  | 830  |
| QY     | 777 | ATCTTTTATGAGGTACAGCCAGCCCTTCTCCAACTTTGTTAGAGTTTACACCTCCATC    | 836  |
| DB     | 831 | CTCTTTTCTGTGGAGTACCAACAGCCCTCTCAACCTCTAGTAGGACCACTCCACT       | 890  |
| QY     | 837 | GGGTACCAACATTTTCAATCAAGCAGCTTTGACACAGATGAGATCAGCTCTGTGTGGAAGA | 896  |
| DB     | 891 | TCGTACTACATTTTATCAACATTTGCGCAGGAATAGATCAGCTCTGTGTGGAAGA       | 950  |
| QY     | 897 | GCCTATCAGTTTTGTAGTGACACCTTGAGACCTTATACAAATATCTTTTGAAGTTTC     | 956  |
| DB     | 951 | GCCCATCAGCTTTGTGGTGACTCACTTGGAGACATACACAGTATCTCTTTGAAGTTTC    | 1010   |





Db 3168 CCAGGAGCCTGGCAATGTGACAGTGTCTCAAGAACTCTATAAACTGGCAGTATTCAGCTA 3227  
Qy 3174 CTATACATTTGGTTTAAACAGCAAGTACTTCAAGTTGGAAATGGGAATAAAAGCAGTGACAT 3233  
Db 3228 CTATACATTTGGCTTAAACAGCAAGTACTTCAAGTTGGAAATGGGAATAAAAGCAGTGATGT 3287  
Qy 3234 CATTGAAGTATACACAGATCAAGACATACCTG 3265  
Db 3288 AATTCATGTCTACAGACCAAGACATTTTCAG 3319

RESULT 2  
CB167019/c  
LOCUS CB167019 761 bp mRNA linear EST 30-JAN-2003  
DEFINITION BTU602600590.R1 CSEQFL04 small intestine Bos taurus cDNA, mRNA  
sequence.

ACCESSION CB167019  
VERSION CB167019.1 GI:28153144  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.

REFERENCE 1 (bases 1 to 761)  
Adelson,D.L. and Gill,C.A.  
Bovine ESTs (Adelson and Gill)  
Unpublished (2003)  
CONTACT: David L. Adelson  
Animal Breeding and Genetics

Texas A&M University  
Animal Science Dept., TAMU-2471, College Station, TX 77843-2471,  
USA  
Tel: 9798452616  
Fax: 9798456970  
Email: david.adelson@tamu.edu.

FEATURES  
source  
1. .761  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/clone\_lib="small intestine"  
/issue type="small intestine"  
/note="Organ: small intestine; Vector: pBluescript SK+;  
Site.1: NotI; Site.2: EcoRI; sequence 5' of the insert  
(5'-NNN...NNNinsert)  
GCGAATTGGAGCTCCACCGCGGTGGCGCGCGCTCGAG. Sequence 3' of  
the inserts (AAGAATTCGATATCAAGCTTATCGATACGCTGACCTCGAG.  
non-normalized library, sequenced 3' with M13R primer."

ORIGIN  
Query Match 8.8%; Score 625.2; DB 6; Length 761;  
Best Local Similarity 90.0%; Pred. No. 3e-137;  
Matches 682; Conservative 0; Mismatches 73; Indels 3; Gaps 1;

Qy 2767 TACACAGTTTATGCTCGGAATAGATCATCATTTAAAAACTTAAATGCTCACTGAAACATCA 2826  
Db 761 TACACAGTTTATGCTCGGACAGATCATCATTTGAAAACTGTTAAATGTAATGTAACAGTAGTACC 702

Qy 2827 TTGGAGTTATCAGATTTGGATTATAATGTTGAATACAGTGTCTTATGTAACAGTAGTACC 2886  
Db 701 TTAGATTTTTCAGATTTGGACCAATGTTGAATATATGCTTATGTAACAGTAGTACC 642

Qy 2887 AGATTTTGGTATGGGAAAAACAGGAACAATATCATTTAGCTTTTCAAAACACAGAGGGAGCA 2946  
Db 641 AGATTTTGGTATGGGAAAAAGAGAACAATATCATCACTTCGGAACACCCGAGGGAGCA 582

Qy 2947 CCAGCGATCTCCCAAGATGTTTATATGCAAACTCAGTCTTTCATCAATAATCTT 3006  
Db 581 CCAAGTGATCTCCCAAGATGTTTATATGCAAACTCAGTCTTTCATCAATAATCTT 522  
Qy 3007 TTCGTGACACCTCTTTCAAAACCTAATGGGATATACAAATATTTACTCTGTTTATTACAGA 3066

Db 521 TTTTGGACACCTCTTTCAAAACCTAATGGGATTTATACAAATATTTATCTGTTTATTACAGA 462  
Qy 3067 AATACTTCAGGTACTCTTTTATGAGAAATTTTACACTCCATGAACCAATGACTTTGAC 3126  
Db 461 AATACTTCAGGTACTCTTTATCCAGAAATTTTACACTCCATGAAGTA---AATGACTTTGAC 405  
Qy 3127 AATATGACTGTATCCACAATATATAGATAAATGACAATATTCAGCTTACTATATACATTTGG 3186  
Db 404 AATATGACCGTATCTGCAATATATAGATAAATGGAATATTCAGCTTACTATATACATTTGG 345  
Qy 3187 TTAACAGCAAGTACTTTCAGTTGGAAATGGGAATAAAGCAGTGCATCATTTGAAGTATAC 3246  
Db 344 GTAACAGCAAGTACTTTCAGTTGGAAATGGGAATAAAGCAGTGGCATAGTTTCAGGTATAT 285  
Qy 3247 ACAGATCAAGACATACCTGAAGGTTTGTGGAAACCTGACTTACGAATCCATTTGCTCA 3306  
Db 284 ACAGATCAAGACATACCTGAAGGAGTTGTGGGAAACCTGACTTACGAGTCCATTTGCTCA 225  
Qy 3307 ACTGCAATAAATGTAAGCTGGGTCCACCGGCTCAACAAACGGTCTAGTCTTCTACTAT 3366  
Db 224 ACTGAGTAAATGTCAGTGGGGCCCTCCGTCTCAACCAATGGTCTAGTCTTCTACTAT 165  
Qy 3367 GTTTCACATGATCTTACAGCAGACTCCTCGCCATGTGAGACCACTCTTGTTCATATAGAG 3426  
Db 164 GTTTCCTGAGCTTCACAGCAGACTCGCGCCACGTCGAGACCACCTCTAGTTACGTATGAG 105  
Qy 3427 AGAAGCATATATTTGATTAATCTGGAATAATACATGATATATATAAATAATCTCTCA 3486  
Db 104 AGAAGCATGTAATTTGATAATCTGGACAAATATCTGACTATATCTTTAAAGTCACTCCA 45  
Qy 3487 TCACAGAAAAAGGATTTCTCTGATACCTATCTGCCCCA 3524  
Db 44 TCACAGAAAAAGGATTTCTCTGACGCATATACGGCCCCA 7

RESULT 3  
BE668393 561 bp mRNA linear EST 25-APR-2001  
LOCUS 157119 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.  
DEFINITION BE668393  
ACCESSION BE668393  
VERSION BE668393.1 GI:10028984  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.

REFERENCE 1 (bases 1 to 561)  
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Casas,E., Wray,J.E., White,J., Cho,J., Fahrnkruug,S.C.,  
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,  
Chitko-Mckown,C.G., Perteau,G., Holt,I., Karamycheva,S., Liang,F.,  
Quackenbush,J. and Keefe,J.W.

Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle

Genome Res. 11 (4), 626-630 (2001)  
21180013  
11282978

COMMENT  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390

Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -minmatch 12 options.

PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCAGTCAGCAGC  
Plate: 76 row: I column: 2  
Seq primer: ATTTAGGTGACACTATAG.

```
FEATURES
source      Location/Qualifiers
1..561
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH108"
/clone_lib="MARC 4BOV"
/notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 20 and day 40
embryos."

ORIGIN
Query Match      6.9%; Score 491; DB 2; Length 561;
Best Local Similarity 93.3%; Pred. No. 2.4e-105;
Matches 525; Conservative 0; Mismatches 35; Indels 3; Gaps 1;

QY 6163 GCAAAAACCGTTTCCCAACATATAAACCATATAATAATAACACAGTAAAGCTGATA 6222
DB 1 GCAAAAACCGTTTCCAAACATAAAACCAT---ATAACATAACAGAGTAAAGCTGATA 57
QY 6223 GCTGAGCTAGTGTTCAGGATTCGGATTATATTAATGCCAGCTATATTTCTGCTTATTTA 6282
DB 58 GCTGATGCTAGTATTCAGGATCAGATTATTAATGCCAGCTATGTTCTGGCTACTTA 117
QY 6283 TGTCCTAAATGAATTTATGCTACTCAAGTCCACTACAGGAACAGTTGGAGATTTTGG 6342
DB 118 TGTCCTAAATGAATTTATGCTACTCAAGTCCACTACAGGAACAGTTGGAGCTTTGG 177
QY 6343 AGAATGGTGGGAACACCGGCAAAACATTAATGCTTAACACAGTGTTTTGA 6402
DB 178 AGAATGGTGGGAACACCGGCAAAACATTAATGCTTAACACAGTGTTTTGA 237
QY 6403 GGACCGATCAGATGCCATCAGTATGCGCCAGAGCAACAAGCCAGTACTGCTTTGGA 6462
DB 238 GGACCGATCAGATGCTATCAATATTTGGCCAGAGCAACAACACAGTACTGCTTTGGA 297
QY 6463 GATATAGTGAATTAACAAGCTTAATGGAGATGTTTCAATAGATTTGGAATTCAGGGATCTG 6522
DB 298 GATATAGTGAATTAACAAGCTTAATGGAGATGTTTCAATAGATTTGGAATTCAGGGATCTG 357
QY 6523 AAAATTTGAAGGACCGGGATTCGATGCTGTTTGGAGATGTTTCAATAGATTTGGAATTCAGGGATCTG 6582
DB 358 AAAATTTGAAGGACCGGGATTCGATGCTGTTTGGAGATGTTTCAATAGATTTGGAATTCAGGGATCTG 417
QY 6583 GACATGGGTTCTCGAGAACAGCGCCCTCTAATTCACCTTCTGGAAGTGGTTTCAGCA 6642
DB 418 GAACACGGGTTCTCGAGAACAGCGCCCTCTAATTCACCTTCTGGAAGTGGTTTCAGCA 477
QY 6643 AGCAGGGCACATGACACCAACCATATGATTTGTTTCACTGCAAGTCTGGAGTTGGAAGAACT 6702
DB 478 AGCAGGGCACATGACACCAACCATATGATTTGTTTCACTGCAAGTCTGGAGTTGGAAGAACT 537
QY 6703 GGAGTTTATTTGCTCTGGACCA 6725
DB 538 GGAGTTTATTTGCTCTGGACCA 560

RESULT 4
BE751055 538 bp mRNA linear EST 25-APR-2001
LOCUS 202768 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION BE751055
ACCESSION BE751055
VERSION BE751055.1 GI:10165047
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 538)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
```

```
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
Chitko-Mckown,C.G., Perte,G., Holt,I., Karamycheva,S., Liang,F.,
Quackenbush,J. and Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
JOURNAL MEDLINE
PUBMED 11282978
COMMENT
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGAAACACGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 42 row: H column: 16
Seq primer: ATTAGTGCACATATAG.
FEATURES
Location/Qualifiers
1..538
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH108"
/clone_lib="MARC 2BOV"
/notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."

ORIGIN
Query Match      6.8%; Score 485.2; DB 2; Length 538;
Best Local Similarity 93.9%; Pred. No. 5.8e-104;
Matches 505; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 6419 ATCAGTATGGCCAGAGGACAAACAGCCAGTTACTGCTTTGGAGATATAGTATTACAA 6478
DB 1 ATCAATATTTGGCCAGAGGACAAACAGCCAGTTACTGCTTTGGAGATATAGTATTACAA 60
QY 6479 AGCTAATGAGGAGATGTTCAATAGATTGGAATCTAGGGATCTGAAATTTGAAGGCATG 6538
DB 61 AACTAATGAGGAGATGTTCAATAGATTGGAATCTAGGGATCTGAAATTTGAAGGCATG 120
QY 6539 GGGATTGCAATGATGTTTCGACAGTGAATCTTACTGCTGGCCAGAGCATGGGGTTCTCTG 6598
DB 121 GGGATTGCAATGATGTTTCGACAGTGAATCTTACTGCTGGCCAGAGCATGGGGTTCTCTG 180
QY 6599 AGAACAAGCCCTCTTAATTTCACTTTGTAAGTTGTTTCGAGCAACGACGACATGACA 6658
DB 181 AGAACAAGCCCTCTCTAGTCCACTTTGTAAGTTGTTTCGAGCAACGACGACATGATA 240
QY 6659 CCACACTTATGATTTGTTTCACTGCAAGTCTGGAAGATCTGGAAGAACTGGAGTTTATTTGCTC 6718
DB 241 CCACACTTATGATTTGTTTCACTGCAAGTCTGGAAGATCTGGAAGAACTGGAGTTTATTTGCTC 300
QY 6719 TGGACCATTTTAAACAAATATAAATGACCATGATTTTGTGGATATATATGACTAGTAG 6778
DB 301 TGGACCATTTTAAACAAATATAAATGATCAGCATTTTGTGGATATATATGACTAGTAG 360
QY 6779 CTGCACTGAGAGTGAAGAAATGTCATGTCAGATCTGGCAGATCTGGCAGATATCTTTTAC 6838
DB 361 CTGCACTGAGAGTGAAGAAATGTCATGTCAGATCTGGCAGATCTGGCAGATATCTTTTAC 420
QY 6839 ACCAGTGCATCTTGGATCTCTTATCAAAATGAAGGAAGTAAATCAGCCCATCTGTTTGTGA 6898
DB 421 ATCAGTGCATCTTGGATCTCTTATCAAGTAAGGAAGTAAATCAGCTCATCTGTTTGTGA 480
QY 6899 ACTATTGACACTTTCAGAGATGGACTCTTTTGGACCCATGGAAGGTGATGTTGAGCT 6956
```

```
Db 481 ACTATTGACACTTCAGAAGTGGACTCTTTGGATGCCATGGAAGTGATGTTGAGCT 538
|||||
BM750287 517 bp mRNA linear EST 04-MAR-2002
K-EST0025803 S5SNU484 Homo sapiens cDNA clone S5SNU484-12-C05 5',
mRNA sequence.
ACCESSION BM750287
VERSION BM750287
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 517)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 12 row: C column: 05
High quality sequence stop: 517.
FEATURES
Location/Qualifiers
1..517
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S5SNU484-12-C05"
/sex="M"
/tissue_type="Stomach"
/cell_type="Epithelial"
/lab_host="SNU-484"
/lab_host="Top10F"
/clone_lib="S5SNU484"
/note="Organ: Stomach; Vector: pTZ19RPI; Site 1: EcoRI;
Site 2: NotI; The poly (A)+ RNA was decapped with tabacco
acid pyrophosphatase (TAP) and ligated with DNA-RNA linker
including EcoRI site by treatment of T4 RNA ligase. The
first strand cDNA was synthesized from oligo dt-selected
mRNA by priming with dt-tailed vector. The dt-tailed
vector was adjusted to have about 60nt. The cDNA vector
was circularized with E. coli DNA ligase after digestion
of EcoRI which site is also included in vector. An RNA
strand converted to a DNA strand by Okayama-Berg method.
The obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation
method."
ORIGIN
Query Match 5.7%; Score 401.8; DB 4; Length 517;
Best Local Similarity 99.5%; Pred. No. 4e-84;
Matches 403; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3718 ATATTGAGAGCTTTCACCATTTACATTATATAGCTTTTGTCTCCGCGAAGACTAGA 3777
|||||
Db 3 ATATAGGAGAGCTTTCACCATTTACATTATATAGCTTTTGTCTCCGCGAAGACTAGA 62
QY 3778 AAAGGACTTGCTTCCAGTATTTCTTTCTTTTACACAGATGAGTCAGTCGCGTTAGCA 3837
|||||
Db 63 AAAGGACTTGCTTCCAGTATTTCTTTCTTTTACACAGATGAGTCAGTCGCGTTAGCA 122
QY 3838 CCTCCACAAATTTGACTTTAATCAACTGCTGACTTCAGACTTTGTATGGCTGAAATGAGC 3897
|||||
Db 123 CCTCCACAAATTTGACTTTAATCAACTGCTGACTTCAGACTTTGTATGGCTGAAATGAGC 182
```

```
QY 3898 CCAAGTCTCTTCCAGSTGGTATTGTTAAAGTATATAGTTTTAAATTCATGAACATGAA 3957
|||||
Db 183 CCAAGTCTCTTCCAGSTGGTATTGTTAAAGTATATAGTTTTAAATTCATGAACATGAA 242
|||||
QY 3958 ACTGACACTATATATTATAAGAAATATATCATCAGATTTAAACATGAAGCCAACTTGTGGA 4017
|||||
Db 243 ACTGACACTATATATTATAAGAAATATATCATCAGATTTAAACATGAAGCCAACTTGTGGA 302
|||||
QY 4018 CTGGAACCACTGACGACCACTCTCTATCCGTGTATCTGCGTTCCACAAAGTTCGAAATGCG 4077
|||||
Db 303 CTGGAACCACTGACGACCACTCTCTATCCGTGTATCTGCGTTCCACAAAGTTCGAAATGCG 362
|||||
QY 4078 AATCAATTTAGTAATGTAGTAAATTCACAAACCCCAAGAAATCAGTT 4122
|||||
Db 363 AATCAATTTAGTAATGTAGTAAATTCACAAACCCCAAGAAATCAGTT 407
|||||
RESULT 6
BB230596 689 bp mRNA linear EST 23-OCT-2001
LOCUS BB230596
DEFINITION musculus cDNA clone A630028F16 3', mRNA sequence.
ACCESSION BB230596
VERSION BB230596
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 689)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
Ohno,M., Okazaki,Y., Okido,T., Saito,K., Sakai,C., Sakai,K.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tanigawa,A., Takahashi,F.,
Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Jul 4, 2000 this sequence version replaced gi:8910509.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yananaka,I.,
Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
```

Db 482 ATAAACTGGCAGTATTCAGCTACTATACATTTTGGCTAAACAGCAAGTACTTTAGTTGGAA 541

Qy 3212 ATGGGAATAAAGCAGTGACATCATTTGAAGTATACACAGATCAAGACATACCTG 3265

Db 542 ATGGGAATAAAGCAGTGATGAATTCATGTTCTACAGACCAAGACATTTTCAAG 595

## 1. .689

```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A630028F16"
/tissue_type="thymus"
/dev_stage="3 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 3 days neonate
thymus"
/notes="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5', GAGAGAGAGATCTCGAGTTAATTAATATCCCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

```

# ORIGIN

| Query Match           | 5.5%; Score 388.4; DB 2; Length 589;                                 |
|-----------------------|--|
| Best Local Similarity | 78.1%; Pred. No. 6.6e-81;  |
| Matches               | 464; Conservative 0; Mismatches 130; Indels 0; Gaps 0;               |
| 2672                  | CTCCCCCTCAAGACTTCTCTGTAAACAGTTGCTCGTGTACGGTGAAGTGTCTCATGCC 2731      |
| dbb                   |  |
| 2                     | CTTTCCTCCCANATTTTCCCGGGAACGTTGTTCTCGGGTCCCGGNATGTTGTCTCATGCC 61      |
| dbb                   |  |
| 2732                  | AAACACCCCTGGAGCCAAATGGAATTTATCTTTATTACACAGTTTATGTCTGGAATAGAT 2791    |
| dbb                   |  |
| 62                    | ACCCCTCTGGAGCCAAATGGGAATTTCTTTATTACCCAGTTTAATGTTGGGATAAG 121         |
| dbb                   |  |
| 2792                  | CATCATTAATAAACTATTAAATGTCACCTGAACATCATTTGGAGTTATCAGATTTGGATTATA 2851 |
| dbb                   |  |
| 122                   | TATCATTAATAAACTATTAAATGCAATGAAGTATCNCGGNATTTATCCGACTTAGACTATC 181    |
| dbb                   |  |
| 2852                  | ATGTTGAATACAGTGCTTATGTAAACAGCTAGCACAGATTTGGTGAATGGAAACAGGAA 2911     |
| dbb                   |  |
| 182                   | ATGCTGACTATAGCGGCTATGTAAACAGCTAGCCCTAGATTTGGTGAATGGAAACAGGAA 241     |
| dbb                   |  |
| 2912                  | GCAATATCATTTAGCTTTCAACACACAGAGGGAGCACCAAGCGATCTCTCCCAAGATGTTT 2971   |
| dbb                   |  |
| 242                   | GCAGCGTCAATTAACCTTCGAAACCCACAGAGGGAGCCAAAGTGATCTCTCCCAAGATGTCC 301   |
| dbb                   |  |
| 2972                  | ATTATGCAAACTCGAGTCTTTCATCAATTAATCTTTTCTGGACACCTCTTCAAAACCTA 3031     |
| dbb                   |  |
| 302                   | CCTATGTGAACCTCGAGTCTTTTATCGAATAATTTCTTTTGGAGCTCTCTCTGTAACACCTA 361   |
| dbb                   |  |
| 3032                  | ATGGGAATATACAATAATTAATCTGTTTATTACAGAAATACTTTCAGGTACTTTTATTCAGA 3091  |
| dbb                   |  |
| 362                   | ATGGAATCATACAATAATTAATCTGTTTATTACAAATAACTTTCAGTACTTTTGTGCAGA 421     |
| dbb                   |  |
| 3092                  | ATTTTACACTCCATGAACCTAAACCAATGACTTTTGACAATATGACTGTATCCAAATTTATAG 3151 |
| dbb                   |  |
| 422                   | ATTTTACACTCTCGAAGTAAACCCAGAGCCTGGCAATGTGACAGTGTTCGAGNATCT 481        |
| dbb                   |  |
| 3152                  | ATAAATCGACAATATTCAGTACTATATACATTTTGGTTAAACAGCAAGTACTTTCAGTTGGAA 3211 |
| dbb                   |  |

```

e mouse tissues.
Location/Qualifiers
1. .605
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A630028F16"
/tissue_type="thymus"
/dev_stage="3 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length
thymus"

```

```

e mouse tissues.
location/Qualifiers
1. .605
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A630028f16"
/tissue_type="thymus"
/dev_stage="3 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 3 days neonate
thymus"

```



QY 3805 TTCTTTTACACAGATGAGTCAGTCCGCTTAGCAGCTCCACAAATTTGACTTTAAATCAAC 3864  
 Db 362 CATTTCTACACGGATGAGTCAGTCCGCTTAGCAGCTCCACAAATTTGACTTTAAATCAAC 421  
 QY 3865 TGTACTTTCAGACTTTGATGCTGAAATGAGCCGAGCTCTCTCCAGGTGTTATTT 3924  
 Db 422 TACACTGCAGATCTCTGTGGCTAAATGGAACCAAGTCTCAGCAATGTTGTTAT 481  
 QY 3925 AAAGTATATAGTTTTAAATTTCAATGAACTGACATATATATATATATAAGAAATATA 3984  
 Db 482 ACAAGATATATCTAAGATTTTATCAAAACGACACTGAAAAATTTCTATCAGATATT 541  
 QY 3985 TCAGGATTTAAACTGAAGCAACTTTGGAAGTGAACCACTGAGCCTACTCTATC 4044  
 Db 542 TCAGGTTCCAACTGAGGCAAGCTCGATGGATTAACCACTTATTTTATC 601  
 QY 4045 CGTGTATCTGCTTCCACCA-AGTTGGAATGCAATCAATTTAGTAATGTAAGAAAT 4103  
 Db 602 AGTGTCTCTGCAATTTACCAAGACTTGGAAAGCAATCAGTTCAATGCTCTCCAGT 661  
 QY 4104 CACAACCCAGAA 4116  
 Db 662 TACAAGCAATGGA 674

RESULT 9  
 A0812654  
 LOCUS  
 DEFINITION HS\_5266\_B1\_H03\_SP6E\_RPC1-11 Human Male BAC Library Homo sapiens  
 genomic clone Plate=842 Col=5 Row=P, genomic survey sequence.  
 ACCESSION A0812654  
 VERSION A0812654.1 GI:5773632  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 484)  
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.  
 Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
 99380589  
 10449764  
 Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPC1-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieterdejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering/bac.htm)  
 or from Research Genetics (info@resgen.com). BAC end web Server:  
 http://www.htsc.washington.edu  
 Plate: 842 row: P column: 5  
 Seq primer: SP6  
 Class: BAC ends  
 High quality sequence stop: 484.

FEATURES  
 source  
 1. .484  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="Plate=842 Col=5 Row=P"  
 /sex="male"  
 /clone\_lib="RPC1-11 Human Male BAC Library"  
 /note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;

Male blood DNA was isolated from one randomly chosen donor  
 and partially digested with a combination of EcoRI and  
 EcoRI Methylase. Size selected DNA was cloned into the  
 pBACe3.6 vector at EcoRI sites"

ORIGIN  
 Query Match 4.2%; Score 301.6; DB 8; Length 484;  
 Best Local Similarity 97.0%; Pred. No. 2.7e-60;  
 Matches 318; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
 QY 4389 AGTTCCTCAGTGTTCCTCCACAAATATTCCTTTCTGATGTTTCAGTCAACTAGTGCACAACTT 4448  
 Db 68 AGTTCCTCAGTGTTCCTCCACAAATATTCCTTTCTGATGTTTCAGTCAACTAGTGCACAACTT 127  
 QY 4449 GACATGATTAAGACCTGACACTATCTCTGGCTACTTTCAAAATTTACAAAATTTACCACTCA 4508  
 Db 128 GACATGATTAAGACCTGACACTATCTCTGGCTACTTTCAAAATTTACAAAATTTACCACTCA 187  
 QY 4509 ACTTCGTCTCAAAATGCAAGAAATGGAATCCGAAGAAATGTTGTAATATCAAAAAT 4568  
 Db 188 ACTTCGTCTCAAAATGCAAGAAATGGAATCCGAAGAAATGTTGTAATATCAAAAAT 247  
 QY 4569 TCAATACCTCTATGAAGCTCACTTAACTGAAGACAGATATATGGAATTAAGAAATTTAG 4628  
 Db 248 TCAATACCTCTATGAAGCTCACTTAACTGAAGACAGATATATGGAATTAAGAAATTTAG 307  
 QY 4629 ATGGTATAGATTCCTCAAGTGGCTGCCAGACCACTATGCTGGCTATGCAATGCTTCAAACTG 4688  
 Db 308 ATGGTATAGATTCCTCAAGTGGCTGCCAGACCACTATGCTGGCTATGCAATGCTTCAAACTG 367  
 QY 4689 GATTTCACAAAATCTCTGCTGGCCCT 4716  
 Db 368 GATTTCAC-AAAATCTCTGCTGGCCGT 394

RESULT 10  
 CE385339  
 LOCUS  
 DEFINITION tigr-gas-dog-17000334364787 Dog Library Canis familiaris genomic,  
 genomic survey sequence.  
 ACCESSION CE385339  
 VERSION CE385339.1 GI:36619903  
 KEYWORDS GSS.  
 SOURCE Canis familiaris (dog)  
 ORGANISM Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 1 (bases 1 to 771)  
 Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,  
 Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and  
 Venter,J.C.  
 The dog genome: survey sequencing and comparative analysis  
 Science 301 (5641), 1898-1903 (2003)  
 22875432  
 14512627  
 Contact: Kirkness EF  
 The Institute for Genomic Research  
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
 Rockville, MD 20850, USA  
 Tel: 301-838-0200  
 Fax: 301-838-0208  
 Email: ekirknes@tigr.org  
 Class: shotgun.  
 Location/Qualifiers  
 1. .771  
 /organism="Canis familiaris"  
 /mol\_type="genomic DNA"  
 /strain="Standard Poodle"  
 /db\_xref="taxon:9615"  
 /clone\_lib="Dog Library"  
 /note="Site 1: BstXI; Libraries were prepared from  
 peripheral blood"

FEATURES  
 source  
 1. .771  
 /organism="Canis familiaris"  
 /mol\_type="genomic DNA"  
 /strain="Standard Poodle"  
 /db\_xref="taxon:9615"  
 /clone\_lib="Dog Library"  
 /note="Site 1: BstXI; Libraries were prepared from  
 peripheral blood"



```
Query Match      3.9%; Score 277; DB 9; Length 771;
Best Local Similarity 88.3%; Pred. No. 2.2e-54; Indels 0; Gaps 0;
Matches 301; Conservative 0; Mismatches 40;

QY 4389 AGTTCACAGTGTCCACAAATATTCCTTTCTGATGTTCACTCAACTAGTGCAACATT 4448
Db 38 AGTTCACAGTGTCCACAAATATTAATTTTCTATGTTTCAGGCCACTAGTGCAACATT 97
QY 4449 GACATGGATAGACCTGACACTATCTCTGGCTACTTTTCAAAATACAAAATTACCACTCA 4508
Db 98 AACATGGATAGACCTGACACTATTTTGGCTACTTTCCAAAATTAACAAGATAACCACTCA 157
QY 4509 ACTTCTGTCTCAAAAATGCAAGAATGGAATCCGAAGAATGTGTTGAATATCAAAAAT 4568
Db 158 ACTCGGGCTCAGAAATGCTGAATGGGATTTCTGAAGAATGTGTTGAATATCAAAAAT 217
QY 4569 TCAATACCTCTATGAAGCTCACTTAATCTGAAGAGACAGTATATGGAATTAAGAAATTTAG 4628
Db 218 TCAATACCTTTATGAAGCTGACCTAACTGAGAGACAGTATGGAATTAAGAAATTTAG 277
QY 4629 ATGTTATAGATTCGAAGTGGTCCAGACCAATGCTGGCTATGGCAATGCTTCAAACATG 4688
Db 278 ATGTTATAGATTCGAAGTGGTCCAGACCAATGCTGGCTATGGCAATGCTTCAAATG 337
QY 4689 GATTCTTACAAAACCTCTCGCTGGCCCTCCAGATGTCCTC 4729
Db 338 GATATCTACAAAACCTCTCTGCTGTAAGTATTAATCTTC 378

RESULT 11
AW081039/c
LOCUS
DEFINITION
  412 bp mRNA linear EST 14-OCT-1999
  xc39f04.x1 NCI CGAP Co20 Homo sapiens cDNA clone IMAGE:2586655 3'
  similar to TR:088488 088488 GLOMERULAR MESANGIAL CELL RECEPTOR
  PROTEIN-TYROSINE PHOSPHATASE PRECURSOR ;, mRNA sequence.
ACCESSION
  AW081039
VERSION
  AW081039.1 GI:6036191
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 412)
AUTHORS
  Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
  Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
  A Comprehensive Collection of Chicken cDNAs
  Curr. Biol. 12 (22), 1965-1969 (2002)
JOURNAL
  MEDLINE
  22335534
PUBMED
  12445392
COMMENT
  Contact: Simon Hubbard
  Department of Biomolecular Sciences
  University of Manchester Institute of Science and Technology
  (UMIST)
  PO Box 88, Manchester, M60 1QD, UK
  Tel: 01612008930
  Fax: 01612360409
  Email: Simon.Hubbard@umist.ac.uk.
  Location/Qualifiers
  1. .746
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:2586655"
  /tissue_type="moderately differentiated adenocarcinoma"
  /lab_host="DH10B"
  /clone_lib="NCI_CGAP Co20"
  /note="Organ: colon; Vector: pCMV-SPORT6; Site 1: SalI;
  Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.
  Normalized to Cot 500. Average insert size 1.1kb.
  Normalized version of NCI_CGAP_Col8. Library constructed
  through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html
  Seq primer: -40UP from Gibco
  High quality sequence stop: 379.
  Location/Qualifiers
  1. .412
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:2586655"
  /tissue_type="moderately differentiated adenocarcinoma"
  /lab_host="DH10B"
  /clone_lib="NCI_CGAP Co20"
  /note="Organ: colon; Vector: pCMV-SPORT6; Site 1: SalI;
  Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.
  Normalized to Cot 500. Average insert size 1.1kb.
  Normalized version of NCI_CGAP_Col8. Library constructed
  through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html
  Seq primer: -40UP from Gibco
  High quality sequence stop: 379.
  Location/Qualifiers
  1. .746
  /organism="Gallus gallus"
  /mol_type="mRNA"
  /strain="White Leghorn, Hisex"
  /db_xref="taxon:9031"
  /clone="CHEST43116"
  /tissue_type="whole embryo"
  /dev_stage="20-21"
  /lab_host="DH10B"
  /clone_lib="CSEQCHN04"
  /note="Organ: whole embryo; Vector: pBluescript II KS(+);
  Site 1: EcoRI; Site 2: NotI; This normalized library was
  constructed from 1 million independent clones. cDNA
  synthesis was initiated using an oligo(dT) primer, using
  methylated C in the first strand synthesis reaction.
  Following this first strand reaction, double-stranded cDNA
  was blunted, ligated to NotI adapters, digested with
  EcoRI, size-selected, and cloned into the NotI and EcoRI
  compatible sites of a custom modified MCS of the
  pBluescript (KS+) vector. The library was normalized in 2
  by Life Technologies."
```

rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bernaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

```
ORIGIN
Query Match      3.0%; Score 215; DB 5; Length 746;
Best Local Similarity 72.5%; Pred. No. 1.2e-39;
Matches 342; Conservative 0; Mismatches 106; Indels 24; Gaps 4;

QY 6655 GACACACACACCTATGTTCTCACTGCAGTGTGGAGTTGGAGAACTGGAGTTTATTT 6714
DB 2 GATATATACCGCATGGTGTCTACTGCAGTGTGGTGT-----GGAAGACAGGTGTATA 57

QY 6715 GCTCTGGACACATTTAAACAACATATAAATGACCATGATTTTGTGGATATATATGACAT 6774
DB 58 TGCTCTGACCATTTAAACGCAACATATAAATGACCATGATTTTGTGGATATATATGACAT 116

QY 6775 GTAGCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 6834
DB 117 GTAGCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 176

QY 6835 TTACACACAGTGCATCTCTGATCTCTTATCAATTAAGGGAAGTAATCAGCCCATCTGTTT 6894
DB 177 TTGATCATGATGCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGAT 236

QY 6895 GTTAATCTATTCAGCACTTCAGAGATGAGTCTCTTTGGACGCCATGGAA--GGTGAATGGA 6953
DB 237 GTAAATTTATTCAGCACTTCAGAGATGAGTCTCTTTGGACGCCATGGAA--GGTGAATGGA 296

QY 6954 GCTTGAATGGGAAGAAACACATATGTAATATTCAGACCAAGGATACA-----7002
DB 297 GCTTGAATGGGAAGAAACACATATGTAATATTCAGACCAAGGATACA-----7002
QY 7003 -----ATTGGAAGATTTTAAATCCAGGGCCAAAGTTACCCCTCATCTTCG 7055
DB 357 GAAAGTCAAGTCATATTTTCTTAAGCAAAAGGCCCAAGCACATCTCCCTATTTGATG 416

QY 7056 AATTGAAATGTGCAACCTTAAAGAAATATCTATGCTTCTCTCACTGTGCTT 7107
DB 417 ATTAAACACAGACACAAATGTTGAACAGCTTTCTTCTATCAATGTGCTT 468
```

```
RESULT 13
LOCUS B66492 629 bp DNA linear GSS 21-JUN-1998
DEFINITION CIT-HSP-2011J12.TF CIT-HSP Homo sapiens genomic clone 2011J12, genomic survey sequence.
ACCESSION B66492
VERSION B66492.1 GI:2640470
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 629)
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,B., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other GSSs: CIT-HSP-2011J12.TF CIT-HSP-2011J12.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html
```

```
Seq primer: M13 Reverse
Class: BAC ends.
Location/Qualifiers
1..629
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7041638"
/db_xref="taxon:9606"
/clone="2011J12"
/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-HSP"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2: HindIII"
```

```
ORIGIN
Query Match      2.7%; Score 190.8; DB 8; Length 629;
Best Local Similarity 90.3%; Pred. No. 6.8e-34;
Matches 204; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 4798 ATTACTGGACCAACATGTTATCTGATTTGATGTCAAATCGGTAGATTAATGATTAAT 4857
DB 542 AGTGTCTCTCTTAATTTTCTTATATTTTCTATAGTATTAATGATTAATTAAT 483

QY 4858 ATATCTTTCATCAAGTCAATCAAGAAATAAACCACATAGAAATTAAGATTTAGAAATA 4917
DB 482 ATATCTTTCATCAAGTCAATCAAGAAATAAACCACATAGAAATTAAGATTTAGAAATA 423

QY 4918 TTCAACAGGTATTTCTGATGATCACTGATTTTACTGGCAACATTAAGTCTGATATGTA 4977
DB 422 TTCAACAGGTATTTCTGATGATCACTGATTTTACTGGCAACATTAAGTCTGATATGTA 363

QY 4978 GAAGGGAAGTCAAGTCTGATGATTTGTTACTTCTTTTGAATCAG 5023
DB 362 GAAGGGAAGTCAAGTCTGATGATTTGTTACTTCTTTTGAATCAG 317
```

```
RESULT 14
LOCUS CR749546 6600 bp mRNA linear HTC 19-AUG-2004
DEFINITION Homo sapiens mRNA; CDNA DKFP686H15164 (from clone DKFP686H15164).
ACCESSION CR749546
VERSION CR749546.1 GI:51476764
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 6600)
AUTHORS Koehler,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Oeanger,A., Fobo,G., Han,M. and Wiemann,S.
CONSTRM The German cDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (17-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical Research Center at the Heinrich-Heine-University, Dueseldorf/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFP686H15164) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFP686H15164
Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.
```

```
FEATURES
source
1..6600
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFP686H15164"
/tissue_type="uterus endothe1, primary cell culture"
```

|   |                       |  |
|---|-----------------------|--|
| /clone_lib="686 (synonym: hlcc3). Vector pSport1_sfi; host  |                       |  |
| DH10B; sites SfiIA + SfiIB"   |                       |  |
| /dev_stage="adult"  |                       |  |
| /note="protein tyrosine phosphatase, receptor type B  |                       |  |
| precursor, differentially spliced"  |                       |  |
| gene  | 1. .6600              |  |
| CDS   | /gene="DKFp686H15164" |  |
|   | 27. .2786             |  |
|   | /gene="DKFp686H15164" |  |
| /codon_start=1  |                       |  |
| /product="hypothetical protein"   |                       |  |
| /protein_id="CAH18347.1"  |                       |  |
| /db_xref="GI:51476765"  |                       |  |
| translation="MEAEFYMVILCLIPRNSEGFQIVHVKQOCLFKNEKVVVGSCL   |                       |  |
| RIQOQWMTNEKLLHVKSALNSRSPSRSAILDCRQAPWTCYDQSGFL  |                       |  |
| EVENASLFLOKGSRRVVKARKYLHSNMKIDVNEKGLVNESLCLQKAGLGAESVR  |                       |  |
| STNSAPPQILITFNAPDGLFLRLNTEAFIRNAENYSQNSRSHPNLHMTGIT   |                       |  |
| DTSWLSTTPSSSTTEETGLAEPERCNFTLAESKASHSVSIQWRIILGSPCNFLIY   |                       |  |
| SSDTLGAALCPTFRIDNTTGCNLOQAQGLTFYFKIISLDEERTVTLQDPLPAPF  |                       |  |
| GVSKEKTSLSLHVMWTPSSGKVTSEYVGLPDENNQKIQGVOIQESTWNEVTFNLT   |                       |  |
| AGSKYNIATVSGGKRSFVSVTNGSTVSPVKDIGISTKANSLIISHSGSGNERY   |                       |  |
| RLMLMDKGLVHGCVVDKRTSYAFHGLTPGYLYNLTWTEAAGLQNTWRKLVRTAPM   |                       |  |
| EVSNLKVTNDGSLTSUKVQRPDPGNVDSINILSHKGTIKESRVLAPWITETHFEL   |                       |  |
| VPEGLYQVTVSCVSELSAQKAVGRTFPDKVANLEANNGMRSLVSWSPAGDWE  |                       |  |
| QYRIILFNDSVLLNITVGBETQYVMDTGLVPGRQYEVIEVSGNLKNSRQQR   |                       |  |
| TVPLAVQLVRKHANETSLSIMQTPVAWEKVIISLADRLLLHKSLSKDAKEFTF   |                       |  |
| TDLVPCRKYMTVTSISGDLKNSSVKGTVPQVTDLDHVNQGMTSLFTWMTQAG  |                       |  |
| DVEFTQVLLIHENVIKNESISSETSRYSFHSLSKGSLSYVVTVTVSGGISRRQVVE  |                       |  |
| GVTVDPKQGVSVNSARSADYLRVSWVHATGDFHYEVTIRKNNPFIQTKAPPSQKTN  |                       |  |
| VYLFs"  |                       |  |
| ORIGIN  |                       |  |
| Query Match   |                       |  |
| Best Local Similarity 2.7%; Score 190.6; DB 3; Length 6600;   |                       |  |
| Matches 476; Conservative 0; Mismatches 374; Indels 21; Gaps 4;   |                       |  |
| QY  | 6005                  | ACGAGAGATAACCGCGCCCAATAAGCAAGAAATCCCTTCCTGCAACATGTTGAAGAGCTTT 6064       |
| Db  | 5442                  | ACCGNAAACTCTTGTCNATAAATAATACAGTTTGAAGGCGATTCATCAGCTAC 5501               |
| QY  | 6065                  | GCACAAACAAACCTTAAAGTTTCAAGAGAATTTTCGGAATTACCAAAATTTCTTCAGG 6124          |
| Db  | 5502                  | AGGCTGGCTCCAACACTCTTATCCAGAGAAATACGAGGAGTTAAAGACGTGGGCGGAA 5561          |
| QY  | 6125                  | ATCTTCTTCACTGATGCTGATCTGCTCTGGAATAGACGAAACCGTTTCCCAACA 6184              |
| Db  | 5562                  | ACCAGTCATGTGACATTCCTTCCCGGAGAAATAGCGGAAAAATCGATACAACAATA 5621            |
| QY  | 6185                  | TAAACCATATAATAATAACAGAGTAAAGCTGATAGCTGACGCTAGTGTTCAGGTT 6244             |
| Db  | 5622                  | TATTGCCCT---ATGATGCCAGCGAGTGAGCTCTCCAATGTAGATGATGATCTTGT 5678            |
| QY  | 6245                  | CGGATTATTAATGCCAGCTATATTCTGGTTATTATTATGTCCAAATGAATTTATGTCTA 6304         |
| Db  | 5679                  | CTGACTACATCAATGCAGCTACATCCCTGGCAACAACCTTCAGAGAGAATACTTGTC 5738           |
| QY  | 6305                  | CTCAGGTCCTACACAGAACAGTTGGAGATTTTGGAGATGTTGGAGATGTTGGAGAACAGG 6364        |
| Db  | 5739                  | CTCAGGAGCGCTTCTCGCACCAGGATGACTTCTGGAAAAATGGTGGGAAACAACG 5798             |
| QY  | 6365                  | CAAAAACATTAGTAATGCTAACACAGTGTTTTGAAGAGCGGATCAGATCGATCAGT 6424            |
| Db  | 5799                  | TTCAACAATCTGTCATGGTGACCCAGTGTGGAGAGGCGGAGTAAGATGTGACCAT 5858             |
| QY  | 6425                  | ATTGGCCAGAGGACAAACAGCCAGTTACTGTCTTTTGGAGATATAGTGAATTACAAAGCTAA 6484      |
| Db  | 5859                  | ACTGGCCAGCGGACAGGATTCCTCTA---CTATGGGACCTCATCCTCGCATGCTCT 5915            |
| QY  | 6485                  | TGGAGATGTTCAATAGATGTGACTATCAGGATCTGAAATTT-----GAAAGC 6535                |
| Db  | 5916                  | CAGAGTCCGCTCCTGCTGAGTGACCATCCGGAGTTTAAAGATATGTGGTGGAGAACAGC 5975         |
| QY  | 6536                  | ATGGGGATTGCATGATGTTTCGACAGTGAATCTTTACTGCTGCGCCAGAGCATGGGGTTC 6595        |
| Db  |                       | 5976 TTGATGTCACACAGACTCATCCGCCACTTTCACTATACGGTGGCCAGACCATGGAGTCC 6035    |
| QY  |                       | 6596 CTGAGAACAGCGCCCTCTTAATTCACTTTGTGAAGTTGGTTTCGAG-----CAAGCAGGG 6649   |
| Db  |                       | 6036 CAGAAACACACCCAGTCTCTGATCCAGTTTGTGAGAACTGTCAAGGACTACATCAACAGAA 6095  |
| QY  |                       | 6650 CACATGACACACACACTATGATTGTTTCACTGCAGTCTGGAGTTTGAAGAACTGGAGTTT 6709   |
| Db  |                       | 6096 GCCCGGGTCTGGGCCCACTGTGTGCACTGCAGTCTGTGGTAGGACTTGGAACTT 6155         |
| QY  |                       | 6710 TTATTGCTCTGGACCAATTAAACAACAATATAAATGACATCATTTTGTGGATATATATG 6769    |
| Db  |                       | 6156 TTATTGCAATTGGACCCGAACTCTCCAGCAGTTAGACTCCAAAGACTCTGTGGACATTTATG 6215 |
| QY  |                       | 6770 GACTAGTAGCTAACTGAGAGTGAAGAACTGATGTCAGTGTGAGAACTCTGGCAGACATATA 6829  |
| Db  |                       | 6216 GAGCAGTGCACACCTTAAGACTTCACAGGTTTCATGTCCTCAGATGTCAGTATG 6275         |
| QY  |                       | 6830 TCTTTTACACCACTGCATCTTCGTGATCTCTT 6860                               |
| Db  |                       | 6276 TCTACTACATCAGTGTGTAAGAGATGTCCT 6306                                 |
| BHL20957  |                       | 547 bp DNA linear GSS 19-JUL-2001  |
| RPCI-24-230D2-TV RPCI-24  |                       | Mus musculus genomic clone RPCI-24-230D2,                                |
| genomic survey sequence.  |                       |  |
| BHL20957  |                       |  |
| BHL20957.1  |                       | GI:14964469  |
| GSS.  |                       |  |
| Mus musculus  |                       |  |
| Mus musculus (house mouse)  |                       |  |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   |                       |  |
| Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  |                       |  |
| 1 (bases 1 to 547)  |                       |  |
| Tsegeye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,   |                       |  |
| Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,  |                       |  |
| Russell, D., de Jong, P. and Fraser, C.M.   |                       |  |
| Mouse BAC End Sequences from Library RPCI-24  |                       |  |
| Unpublished (1999)  |                       |  |
| Other GSSs: RPCI-24-230D2.TJ  |                       |  |
| Contact: Shaying Zhao   |                       |  |
| Department of Eukaryotic Genomics   |                       |  |
| The Institute for Genomic Research  |                       |  |
| 9712 Medical Center Dr., Rockville, MD 20850, USA   |                       |  |
| Tel: 301 838 0200   |                       |  |
| Fax: 301 838 0208   |                       |  |
| Email: szhao@tigr.org   |                       |  |
| Clones are derived from the mouse BAC library RPCI-24. For BAC  |                       |  |
| library availability, please contact Pieter de Jong   |                       |  |
| (pdejong@mail.cho.org). Clones may be purchased from BACPAC   |                       |  |
| Resources ( <a href="http://www.choi.org/bacpac/orderingframe.htm">http://www.choi.org/bacpac/orderingframe.htm</a> ). BAC end                      |                       |  |
| page: <a href="http://www.tigr.org/tdb/bac_ends/mouse/mouse_bac_end_intro.html">http://www.tigr.org/tdb/bac_ends/mouse/mouse_bac_end_intro.html</a> |                       |  |
| Plate: 230 row: D column: 2   |                       |  |
| Seq primer: T7  |                       |  |
| Class: BAC ends.  |                       |  |
| Location/Qualifiers   |                       |  |
| 1. .547   |                       |  |
| /organism="Mus musculus"  |                       |  |
| /mol_type="genomic DNA"   |                       |  |
| /strain="C57BL/6J"  |                       |  |
| /db_xref="taxon:10090"  |                       |  |
| /clone="RPCI-24-230D2"  |                       |  |
| /sex="Male"   |                       |  |
| /cell_type="Spleen/Brain"   |                       |  |
| /clone_lib="RPCI-24"  |                       |  |
| /note="Vector: pTARBAC1; Site 1: BamH1; Site 2: BamH1;  |                       |  |
| RPCI-24 Mouse BAC Library produced by Pieter de Jong. The   |                       |  |
| library was cloned in the pTARBAC1 cloning vector at the  |                       |  |
| BamH1 sites using MboI partially digested male C57BL/6J   |                       |  |
| DNA."   |                       |  |

ORIGIN

Query Match 2.5%; Score 180.6; DB 8; Length 547;  
Best Local Similarity 87.2%; Pred. No. 1.8e-31;  
Matches 198; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

|    |      |   |      |
|----|------|---|------|
| Qy | 5493 | AGCTCAGCATGATGGAAATGTACAAAGTGGTATGTCATATTTTAATAAGCAAGGCC      | 5552 |
| Db | 444  | AGCTCAGCAAGACGGAAATGTGACAAAGTGGTATGTCATATTTTAACAAAGCAAGGCC    | 385  |
| Qy | 5553 | ATATTTTACAAATGAAGGCTTTCCTAACCCCTCCATGTACACAGAGAAAGCAAAAGTTTAG | 5612 |
| Db | 384  | ATATTTTACAAACGAAGGATTCCTTAATCCCCGTGTATAGAGGAAAGACTTAAGTTTCA   | 325  |
| Qy | 5613 | TGCCAATGAAGAAATCTACATCATAGGTGCTGATAATGCATGCATGATTCCTGGCAATGA  | 5672 |
| Db | 324  | CGGTAATGAAGAAATCTATGTGATAGGTGCTGATAATGCCTGCATGATCCCTGGAATGA   | 265  |
| Qy | 5673 | AGACAAATTTGCAATGGACCCTGAAACCAAAAAGCAATATTAT                   | 5719 |
| Db | 264  | GGAGAAATTTGCAATGGACCCTGAAACCAAAAAGCAATATTAT                   | 218  |

Search completed: July 2, 2005, 21:12:32  
Job time : 32719 secs

***THIS PAGE BLANK (USPTO)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 14:40:39 ; Search time 238 Seconds  
(without alignments)  
3722.975 Million cell updates/sec

Title: US-10-673-885-2  
Perfect score: 11985  
Sequence: 1 MDPLIFLLFLLFCTSETQVD.....MDSLDMEGDVLEWEETM 2291

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score   | Query Match | Length | DB ID | Description        |
|------------|---------|-------------|--------|-------|--------------------|
| 1          | 11985   | 100.0       | 2291   | 6     | ABP60057 Human pro |
| 2          | 11915   | 99.4        | 2299   | 8     | ADH41615 Novel hum |
| 3          | 11905   | 99.3        | 2299   | 6     | Aae37320 Human rec |
| 4          | 11807.5 | 98.5        | 2300   | 5     | Aao18738 Human NOV |
| 5          | 11807.5 | 98.5        | 2300   | 8     | Adh41625 Novel hum |
| 6          | 11797   | 98.4        | 2299   | 6     | Aag79724 Human KPP |
| 7          | 11191   | 93.4        | 2281   | 5     | Aao18736 Human NOV |
| 8          | 11191   | 93.4        | 2281   | 8     | Adh41619 Novel hum |
| 9          | 11191   | 93.4        | 2281   | 8     | Adh41613 Novel hum |
| 10         | 10536   | 87.9        | 2301   | 6     | ABP60058 Human pro |
| 11         | 9874    | 82.4        | 1959   | 8     | Aae37322 Human rec |
| 12         | 4661    | 38.9        | 909    | 8     | Adh41621 Novel hum |
| 13         | 4488    | 37.4        | 855    | 5     | Aao18737 Human NOV |
| 14         | 4482    | 37.4        | 851    | 8     | Adh41623 Novel hum |
| 15         | 3804    | 31.7        | 740    | 8     | Adh41617 Novel hum |
| 16         | 1713.5  | 14.3        | 322    | 2     | Aae60877 Product o |
| 17         | 1539.5  | 12.8        | 330    | 6     | Aae37321 Human rec |
| 18         | 1327    | 11.1        | 2037   | 4     | ABP71928 Human pro |
| 19         | 1284.5  | 10.7        | 459    | 8     | ADP30331 Human sec |
| 20         | 1188    | 9.9         | 1912   | 8     | ADR40183 Human pro |
| 21         | 1160    | 9.7         | 1767   | 4     | ABB62726 Drosophil |
| 22         | 1158.5  | 9.7         | 1907   | 4     | Aau14143 Human nov |
| 23         | 1155.5  | 9.6         | 1897   | 3     | Aay81785 Human pro |
| 24         | 1155.5  | 9.6         | 1897   | 3     | Aay56100 LAR tyros |
| 25         | 1155.5  | 9.6         | 1897   | 3     | Aab19712 Human pro |

|    |        |     |      |   |          |                    |
|----|--------|-----|------|---|----------|--------------------|
| 26 | 1155.5 | 9.6 | 1897 | 7 | ADD18740 | Adi18740 Human dis |
| 27 | 1155.5 | 9.6 | 1897 | 8 | ADJ33670 | Adj33670 Human leu |
| 28 | 1155.5 | 9.6 | 1907 | 8 | ADP18674 | Adp18674 Human pro |
| 29 | 1154   | 9.6 | 1911 | 2 | AAW71726 | Aaw71726 Human PTP |
| 30 | 1154   | 9.6 | 1911 | 2 | AAW27225 | Aaw27225 Human pro |
| 31 | 1154   | 9.6 | 1911 | 2 | AAW94027 | Aaw94027 Human pro |
| 32 | 1154   | 9.6 | 1911 | 4 | AAU01459 | Aau01459 Human pro |
| 33 | 1142.5 | 9.5 | 1948 | 7 | ADD18742 | Add18742 Human dis |
| 34 | 1141.5 | 9.5 | 1647 | 7 | ABBS8428 | Abbs8428 Drosophil |
| 35 | 1141.5 | 9.5 | 1949 | 7 | ABBS7117 | Abbs7117 Human pro |
| 36 | 1141.5 | 9.5 | 1949 | 7 | ADBS7121 | Adbs7121 Human pro |
| 37 | 1141.5 | 9.5 | 1949 | 7 | ADD47019 | Add47019 Human pro |
| 38 | 1141.5 | 9.5 | 1949 | 7 | ADD47015 | Add47015 Human pro |
| 39 | 1126   | 9.4 | 1998 | 7 | AAO24268 | Aao24268 Murine va |
| 40 | 1109.5 | 9.3 | 1904 | 5 | ABBS7100 | Abbs7100 Mouse isc |
| 41 | 1109   | 9.3 | 1997 | 5 | AAE20278 | Aae20278 Human lun |
| 42 | 1108   | 9.2 | 1997 | 7 | ABRS7179 | Abrs7179 Human PTP |
| 43 | 1108   | 9.2 | 1997 | 7 | ADJ70329 | Adj70329 Human hea |
| 44 | 1107   | 9.2 | 1997 | 3 | AAB19774 | Aab19774 Human pro |
| 45 | 1107   | 9.2 | 1997 | 4 | AAM78821 | Aam78821 Human pro |

## ALIGNMENTS

RESULT 1  
ABP60057  
ID ABP60057 standard; protein; 2291 AA.

XX AC ABP60057;

XX DT 21-MAR-2003 (first entry)

XX DE Human phosphatase protein SEQ ID #2.

XX KW Human; phosphatase; enzyme; chromosome 12; colon adenocarcinoma; placenta; gene therapy.

XX OS Homo sapiens.

XX PN WO200279452-A2.

XX PD 10-OCT-2002.

XX PF 02-APR-2002; 2002WO-US009992.

XX PR 02-APR-2001; 2001US-00822871.

XX PA (PEKE ) PE CORP NY.

XX PI Webster M, Wei M, Di Francesco V, Beasley EM;

XX DR WPI: 2003-046809/04.

XX DR N-PSDB; AB223703.

XX PT New human phosphatase peptides and nucleic acids encoding the peptides, useful as models in developing human therapeutic targets, in identifying therapeutic proteins or modulators of protein activity, and in gene therapy.

XX PS Claim 1a; Fig 2; 220pp; English.

XX CC The invention relates to an isolated human phosphatase peptide. The human phosphatase peptides and nucleic acids are useful as models for the development of human therapeutic targets and in the identification of therapeutic proteins. They also serve as targets for the development of human therapeutic agents that modulate phosphatase activity in cells and tissues that express the phosphatase. Experimental data indicates that the cDNA is expressed in colon adenocarcinoma and placenta. The proteins can be used to raise or to elicit another immune response, as a reagent in assays designed to determine the levels of the protein in biological fluids, as markers for tissue in which the corresponding protein is preferentially expressed, in the identification of modulators of the

CC peptides, and in pharmacogenomic analysis. The polypeptides and  
CC polynucleotides may be used in gene therapy, and as antisense  
CC constructs to control phosphatase gene expression in cells, tissue and  
CC organisms. The gene encoding the phosphatase of the invention is located  
CC on human chromosome 12. The current sequence represents the human  
CC phosphatase protein of the invention  
XX

SQ Sequence 2291 AA;

Query Match 100.0%; Score 11985; DB 6; Length 2291;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPLIFLLFLFTGTSTQVDVSNVPGTRYDITISSITTYTSPVTRIIVNTVPGPPV 60  
DB 1 MDPLIFLLFLFTGTSTQVDVSNVPGTRYDITISSITTYTSPVTRIIVNTVPGPPV 60  
QY 61 FLAGERVGSGAILLSWNTTPNPNRGRIISIVYKVCPPMNVQVYTVQVRKPDLSLEVLITN 120  
DB 61 FLAGERVGSGAILLSWNTTPNPNRGRIISIVYKVCPPMNVQVYTVQVRKPDLSLEVLITN 120  
QY 121 LNPGETTYEIKVAENSAGIGVSDPFLPQTASAPGKVNLTVYANASAVKLIWYLPKQ 180  
DB 121 LNPGETTYEIKVAENSAGIGVSDPFLPQTASAPGKVNLTVYANASAVKLIWYLPKQ 180  
QY 181 PNGKITSPKISVGHARSIGVWVDVSRVEDILTGLKPECNENSESFLWSTASPSPLGRV 240  
DB 181 PNGKITSPKISVGHARSIGVWVDVSRVEDILTGLKPECNENSESFLWSTASPSPLGRV 240  
QY 241 TTPSRTHSSSTLTQNEISSVWKEPISFVVTHLRPYTYTLFEVSAAATEAGYIDSTIVRT 300  
DB 241 TTPSRTHSSSTLTQNEISSVWKEPISFVVTHLRPYTYTLFEVSAAATEAGYIDSTIVRT 300  
QY 301 PSVPPEGPQNCVTGNITGKSPSILMDPPTIVTGKFSYRVELYGSGRILDNSTKDKPA 360  
DB 301 PSVPPEGPQNCVTGNITGKSPSILMDPPTIVTGKFSYRVELYGSGRILDNSTKDKPA 360  
QY 361 FTNLTPTWVDVYIAAETAGTGPKSNISVFPPDVPDVGAVFDLQLAESTVQRIITWKKP 420  
DB 361 FTNLTPTWVDVYIAAETAGTGPKSNISVFPPDVPDVGAVFDLQLAESTVQRIITWKKP 420  
QY 421 RQPNGLIINQYRVKVLVPETGIIENLTLLGNNEYINDPMAPEIIVNVEPMVGLYEGSAEM 480  
DB 421 RQPNGLIINQYRVKVLVPETGIIENLTLLGNNEYINDPMAPEIIVNVEPMVGLYEGSAEM 480  
QY 481 SSDLHSLATFIYNHSDPKNPPARNRAEDQTSVVTTRNQYITDIAEQLSYVIRRLVPPT 540  
DB 481 SSDLHSLATFIYNHSDPKNPPARNRAEDQTSVVTTRNQYITDIAEQLSYVIRRLVPPT 540  
QY 541 EHMISVSAFTIMEGPGPTVLSVTRQOVPSIKIINYNKNISSSILLYWDPPPEYNGKIT 600  
DB 541 EHMISVSAFTIMEGPGPTVLSVTRQOVPSIKIINYNKNISSSILLYWDPPPEYNGKIT 600  
QY 601 HVTIYAMELDTNRAFOITIDNSFLITGLKXYTKYKWRVAASTHDESSISENDIFVRT 660  
DB 601 HVTIYAMELDTNRAFOITIDNSFLITGLKXYTKYKWRVAASTHDESSISENDIFVRT 660  
QY 661 SEDEPESSQDVEIVDVTADEIRLKWSPPEKPNGLIIIAEVLVKNIDTLYMKNSTTDII 720  
DB 661 SEDEPESSQDVEIVDVTADEIRLKWSPPEKPNGLIIIAEVLVKNIDTLYMKNSTTDII 720  
QY 721 LRNLRHPTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPDSPAENITYKNISSGETELSP 780  
DB 721 LRNLRHPTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPDSPAENITYKNISSGETELSP 780  
QY 781 LPSSPNGLIIKKYTIYLRSGNGNEERTINTTSLTQNIKVLKXYQYIIIEVSASTLKGEV 840  
DB 781 LPSSPNGLIIKKYTIYLRSGNGNEERTINTTSLTQNIKVLKXYQYIIIEVSASTLKGEV 840  
QY 841 RSAPISILTEEDAPDPPQDFSVKQLSGVTVKLSWOPPLEPNGLIILYYTVVWNRSSLKT 900  
DB 841 RSAPISILTEEDAPDPPQDFSVKQLSGVTVKLSWOPPLEPNGLIILYYTVVWNRSSLKT 900

QY 901 INVTETSLSDLDYNNVEISAYVTASTRFGDKTGSNIISFQTPEGASDPDPKDYVYANL 960  
DB 901 INVTETSLSDLDYNNVEISAYVTASTRFGDKTGSNIISFQTPEGASDPDPKDYVYANL 960  
QY 961 SSSSIILFWTPPSKNGIIQYYSVYRNTSGTFMQNFTLHETLNDPDMVTSTIIDKLTII 1020  
DB 961 SSSSIILFWTPPSKNGIIQYYSVYRNTSGTFMQNFTLHETLNDPDMVTSTIIDKLTII 1020  
QY 1021 PSYTFMLTASTSVGNGKSSDIIIEVYTDODIPEGFVGNLTYESISSSTAINVSWPPAPQ 1080  
DB 1021 PSYTFMLTASTSVGNGKSSDIIIEVYTDODIPEGFVGNLTYESISSSTAINVSWPPAPQ 1080  
QY 1081 NGLVYVYVSLIILOQTPRHRVPLVYERSIYFDNLEKTYDYLKILTPSTEKGSFTYTAQ 1140  
DB 1081 NGLVYVYVSLIILOQTPRHRVPLVYERSIYFDNLEKTYDYLKILTPSTEKGSFTYTAQ 1140  
QY 1141 LYIKTEEDVPETSPIINTFKNLSSTVLLSWDPPVKPNGAIIISYDLTLQGPNEYSFITS 1200  
DB 1141 LYIKTEEDVPETSPIINTFKNLSSTVLLSWDPPVKPNGAIIISYDLTLQGPNEYSFITS 1200  
QY 1201 DNYIILELSPTLYSFFAAARTRKGLGPSSILFFYTDDESVP LAPQNLTLINCTSDFW 1260  
DB 1201 DNYIILELSPTLYSFFAAARTRKGLGPSSILFFYTDDESVP LAPQNLTLINCTSDFW 1260  
QY 1261 LKWSPSPLPGGIVKYVSPKIHHEHTDTIYKNIISGFKTEAKLVGLEPVSTYSIRVSAFTK 1320  
DB 1261 LKWSPSPLPGGIVKYVSPKIHHEHTDTIYKNIISGFKTEAKLVGLEPVSTYSIRVSAFTK 1320  
QY 1321 VGNGQFQSNVVKFTTQESVPDVVQNMOCWATSQSVLVKWDPPKANGIIITQYMTVVERN 1380  
DB 1321 VGNGQFQSNVVKFTTQESVPDVVQNMOCWATSQSVLVKWDPPKANGIIITQYMTVVERN 1380  
QY 1381 STKVSQDQHMVTFIKLANTSVYKVRASTSAGEDESTCHVSTLTPETVPSVPTNIAPSD 1440  
DB 1381 STKVSQDQHMVTFIKLANTSVYKVRASTSAGEDESTCHVSTLTPETVPSVPTNIAPSD 1440  
QY 1441 VQSTSATLTWIRPDITLGYFQNYKITTLQRAQCKEWESEECVEYQKIQYLYEAHLTEB 1500  
DB 1441 VQSTSATLTWIRPDITLGYFQNYKITTLQRAQCKEWESEECVEYQKIQYLYEAHLTEB 1500  
QY 1501 VYGLKKFWRYPQVAASNAGYGNASNMISTKTLPGPDGPPENHVHVSATSPSISWS 1560  
DB 1501 VYGLKKFWRYPQVAASNAGYGNASNMISTKTLPGPDGPPENHVHVSATSPSISWS 1560  
QY 1561 EBAVITGPTCYLIDVKSVDNDEFNIFSKSNEENTKIRIKOLEIFTRYSVWITAFNGNIS 1620  
DB 1561 EBAVITGPTCYLIDVKSVDNDEFNIFSKSNEENTKIRIKOLEIFTRYSVWITAFNGNIS 1620  
QY 1621 AAYVKGKSAEMIIVTTLESAPKDPNNMTFQKIPDEVTKFQTLFPLPPSQPNNGNIQVYQAL 1680  
DB 1621 AAYVKGKSAEMIIVTTLESAPKDPNNMTFQKIPDEVTKFQTLFPLPPSQPNNGNIQVYQAL 1680  
QY 1681 VYREDDPTAVQIHNSIIQKNTTFVIAMLEGLKGGHTNISVYAVNSAGAGPKVPMRITM 1740  
DB 1681 VYREDDPTAVQIHNSIIQKNTTFVIAMLEGLKGGHTNISVYAVNSAGAGPKVPMRITM 1740  
QY 1741 DIKAPARKTKPTPIYDATGKLLVSTTTITRMPICYSDHDPKKNQVOLATETGAQHD 1800  
DB 1741 DIKAPARKTKPTPIYDATGKLLVSTTTITRMPICYSDHDPKKNQVOLATETGAQHD 1800  
QY 1801 GNVTKWYDAYFNKARPYFTNEGPNPPCTEGTKFSGNEEYIIIGADNACMIPGNEDKIC 1860  
DB 1801 GNVTKWYDAYFNKARPYFTNEGPNPPCTEGTKFSGNEEYIIIGADNACMIPGNEDKIC 1860  
QY 1861 NGPLKPKQYLPKFRATNIMGQFTSDYSDDPVKTLGEGLSERTVEIILSVTICILSIIILL 1920  
DB 1861 NGPLKPKQYLPKFRATNIMGQFTSDYSDDPVKTLGEGLSERTVEIILSVTICILSIIILL 1920  
QY 1921 GTAIFAPARIROKQEGGTYSPQDAEIIDTKLQDOLITVADLEKDLRLTRPISKKSFL 1980  
DB 1921 GTAIFAPARIROKQEGGTYSPQDAEIIDTKLQDOLITVADLEKDLRLTRPISKKSFL 1980  
QY 1981 QHVEBLCTNNNLKFOEEFSELPKFLQDLSSTDADLPWNRKRNRPNIKFPYNNNNRVKLI 2040



Db 1981 QHVEELCTNNLKFQSEFSELPKFLQDLSSDADLPWNRKAKRFPNKNYNNNNRKLIA 2040  
QY 2041 DASVPGSDYINASYISGILCPNEFIATQGPLCTGVDFWRMVMWETRAKTLVMLTQCFEKG 2100  
Db 2041 DASVPGSDYINASYISGILCPNEFIATQGPLCTGVDFWRMVMWETRAKTLVMLTQCFEKG 2100  
QY 2101 RIRCHOYWPEDNKPVTVPFGDIVITKLMEDVQIDMTIRDLKIERHGDGMTVRCNFTWAPPE 2160  
Db 2101 RIRCHOYWPEDNKPVTVPFGDIVITKLMEDVQIDMTIRDLKIERHGDGMTVRCNFTWAPPE 2160  
QY 2161 HGVPENSAPLIHFVKLVARSRAHDTTPMLVHCSAGVGRGVFTALDHLTOHINDHDFVDI 2220  
Db 2161 HGVPENSAPLIHFVKLVARSRAHDTTPMLVHCSAGVGRGVFTALDHLTOHINDHDFVDI 2220  
QY 2221 YGLVABLRSERMCWQNLQAYIFLHQCILDLSSNKGSNQPICFVNSALQKQMSLDAMEG 2280  
Db 2221 YGLVABLRSERMCWQNLQAYIFLHQCILDLSSNKGSNQPICFVNSALQKQMSLDAMEG 2280  
QY 2281 DVELEWEEETM 2291  
Db 2281 DVELEWEEETM 2291  
RESULT 2  
ADH41615  
ID ADH41615 standard; protein; 2299 AA.  
AC ADH41615;  
XX  
DT 25-MAR-2004 (first entry)  
DE  
DE Novel human protein NOV15b.  
XX  
KW cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;  
KW antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;  
KW antiparkinsonian; antiasthmatic; antiinfertility; cardiomyopathy;  
KW atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;  
KW multiple sclerosis; graft-versus-host disease; Alzheimer's disease;  
KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;  
KW tissue typing; preventive medicine; pharmacogenomic; vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO2003102159-A2.  
XX  
PD 11-DEC-2003.  
XX  
PF 04-JUN-2003; 2003WO-US017573.  
XX  
PR 04-JUN-2002; 2002US-0385490P.  
PR 04-JUN-2002; 2002US-0385615P.  
PR 04-JUN-2002; 2002US-0385755P.  
PR 05-JUN-2002; 2002US-0386041P.  
PR 06-JUN-2002; 2002US-0386355P.  
PR 06-JUN-2002; 2002US-0386357P.  
PR 06-JUN-2002; 2002US-0386447P.  
PR 06-JUN-2002; 2002US-0386459P.  
PR 06-JUN-2002; 2002US-0386465P.  
PR 06-JUN-2002; 2002US-0386864P.  
PR 07-JUN-2002; 2002US-0386701P.  
PR 07-JUN-2002; 2002US-0386796P.  
PR 07-JUN-2002; 2002US-0386931P.  
PR 07-JUN-2002; 2002US-0387078P.  
PR 07-JUN-2002; 2002US-0387081P.  
PR 07-JUN-2002; 2002US-0387083P.  
PR 10-JUN-2002; 2002US-0387429P.  
PR 10-JUN-2002; 2002US-0387540P.  
PR 10-JUN-2002; 2002US-0387866P.  
PR 11-JUN-2002; 2002US-0387806P.  
PR 11-JUN-2002; 2002US-0387610P.  
PR 11-JUN-2002; 2002US-0387659P.  
PR 11-JUN-2002; 2002US-0387668P.

PR 11-JUN-2002; 2002US-0387696P.  
PR 11-JUN-2002; 2002US-0387859P.  
PR 12-JUN-2002; 2002US-0387934P.  
PR 12-JUN-2002; 2002US-0387960P.  
PR 12-JUN-2002; 2002US-0388022P.  
PR 12-JUN-2002; 2002US-0388096P.  
PR 12-JUN-2002; 2002US-0388432P.  
PR 12-JUN-2002; 2002US-0388439P.  
PR 12-JUN-2002; 2002US-0388479P.  
PR 13-JUN-2002; 2002US-0389123P.  
PR 14-JUN-2002; 2002US-0389120P.  
PR 14-JUN-2002; 2002US-0389146P.  
PR 17-JUN-2002; 2002US-0389742P.  
PR 18-JUN-2002; 2002US-0389604P.  
PR 18-JUN-2002; 2002US-0389884P.  
PR 19-JUN-2002; 2002US-0390006P.  
PR 19-JUN-2002; 2002US-0390144P.  
PR 19-JUN-2002; 2002US-0390209P.  
PR 25-JUN-2002; 2002US-0391726P.  
PR 06-AUG-2002; 2002US-0401628P.  
PR 09-AUG-2002; 2002US-0402268P.  
PR 12-AUG-2002; 2002US-0402822P.  
PR 13-AUG-2002; 2002US-0403459P.  
PR 15-AUG-2002; 2002US-0403617P.  
PR 15-AUG-2002; 2002US-0403732P.  
PR 26-AUG-2002; 2002US-0406182P.  
PR 12-SEP-2002; 2002US-0410085P.  
PR 13-SEP-2002; 2002US-0410505P.  
PR 23-SEP-2002; 2002US-0412955P.  
PR 30-SEP-2002; 2002US-0415195P.  
PR 23-OCT-2002; 2002US-0420627P.  
PR 23-OCT-2002; 2002US-0420718P.  
PR 24-OCT-2002; 2002US-0420852P.  
PR 31-OCT-2002; 2002US-0422750P.  
PR 01-NOV-2002; 2002US-0423095P.  
05-NOV-2002; 2002US-0423748P.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;  
PI Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;  
PI Ellerman K, Rittenberg S, Gangolli EA, Gerlach VL, Gorman L;  
PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;  
PI Khrantsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;  
PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CB;  
PI Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;  
PI Rieger DK, Rothenberg ME, Shenoy SG, Shinkete RA, Smithson G;  
PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CM, Voss EZ;  
PI Wolenc AR, Zhong M, Zhong H;  
XX  
XX WPI; 2004-053467/05.  
DR N-PSDB; ADH41614.  
XX  
XX New NOVX polypeptides and nucleic acid molecules useful for preventing or  
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in  
PT pharmacogenomics.  
XX  
XX Claim 2; SEQ ID NO 168; 1503pp; English.  
PS  
XX The invention relates to 566 new isolated human polypeptides and their  
CC encoding genes, sequences that are at least 95% identical to these or  
CC sequences comprising one or more conservative substitutions in these. The  
CC polypeptide, polynucleotide and antibodies against the polypeptides are  
CC useful in diagnosing, treating or preventing NOVX-associated disorders,  
CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,  
CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,  
CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.  
CC The nucleic acids are further used as hybridization probes, in chromosome  
CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The  
CC polypeptides are also useful as vaccines. This sequence represents an  
CC example of the polypeptide of the invention.  
XX  
SQ Sequence 2299 AA;

|                       |      |   |              |            |              |                    |
|-----------------------|------|---|--------------|------------|--------------|--------------------|
| Query Match           |      | 99.4%;  | Score 11915; | DB 8;      | Length 2299; |                    |
| Best Local Similarity |      | 99.2%;  | Pred. No. 0; |            |              |                    |
| Matches 2282;         |      | Conservative  | 3;           | Mismatches | 5;           | Indels 10; Gaps 2; |
| QY                    | 1    | MDPLIIFLLFGTSETQDVSNVPGTRYDITISSITTYTSPVTRIVTPNVTKPGPPV           | 60           |            |              |                    |
| DB                    | 1    | MDPLIIFLLFGTSETQDVSNVPGTRYDITISSITTYTSPVTRIVTPNVTKPGPPV           | 60           |            |              |                    |
| QY                    | 61   | FLAGERVGSAGILLSNWTPPNNGRIISIVYKVEKVCPPMOTVYTOVRKPSDLEVLITN        | 120          |            |              |                    |
| DB                    | 61   | FLAGERVGSAGILLSNWTPPNNGRIISIVYKVEKVCPPMOTVYTOVRKPSDLEVLITN        | 120          |            |              |                    |
| QY                    | 121  | LNPGETTYEIKVAENAGIGVSDPELFOFASAPCKVNLTVYENASAVKLIWLPQ             | 180          |            |              |                    |
| DB                    | 121  | LNPGETTYEIKVAENAGIGVSDPELFOFASAPCKVNLTVYENASAVKLIWLPQ             | 180          |            |              |                    |
| QY                    | 181  | PNGKITSPKISVKHARSAGIIVKDVSIKVSILWDPTITVTGKFSYRVELYGPGRILDNSTDKLKA | 360          |            |              |                    |
| DB                    | 181  | PNGKITSPKISVKHARSAGIIVKDVSIKVSILWDPTITVTGKFSYRVELYGPGRILDNSTDKLKA | 360          |            |              |                    |
| QY                    | 241  | TPPSRTTHSSSLTQNEISVWKEPISFVWTHLRPYTTLFEVSAVTTAGYIDSTIVRT          | 300          |            |              |                    |
| DB                    | 241  | TPPSRTTHSSSLTQNEISVWKEPISFVWTHLRPYTTLFEVSAVTTAGYIDSTIVRT          | 300          |            |              |                    |
| QY                    | 301  | PSVPEGPQNCVGTNITGKFSILWDPTITVTGKFSYRVELYGPGRILDNSTDKLKA           | 360          |            |              |                    |
| DB                    | 301  | PSVPEGPQNCVGTNITGKFSILWDPTITVTGKFSYRVELYGPGRILDNSTDKLKA           | 360          |            |              |                    |
| QY                    | 361  | FTNLTPFTMYDVYIAAETSAAGTGPKNISVFTPPDVPDVGAVFDQLAEVESTQVRIWKKP      | 420          |            |              |                    |
| DB                    | 361  | FTNLTPFTMYDVYIAAETSAAGTGPKNISVFTPPDVPDVGAVFDQLAEVESTQVRIWKKP      | 420          |            |              |                    |
| QY                    | 421  | RQNGIINOYRVKVLVPETGIIENLTLLTGNNEYINDPMAPEIVNIPEPMVGLYEGSAEM       | 480          |            |              |                    |
| DB                    | 421  | RQNGIINOYRVKVLVPETGIIENLTLLTGNNEYINDPMAPEIVNIPEPMVGLYEGSAEM       | 480          |            |              |                    |
| QY                    | 481  | SSDLHSLATFIYNHSDPKNPPARNRAEDQTSVPTTRNOYITDIAEQLSYVIRLVPET         | 540          |            |              |                    |
| DB                    | 481  | SSDLHSLATFIYNHSDPKNPPARNRAEDQTSVPTTRNOYITDIAEQLSYVIRLVPET         | 540          |            |              |                    |
| QY                    | 541  | EHMISVSAFTIMEGEPPTVLSVTRQOVPSIKIINYKNISSSILLYWDPPPEYNGKIT         | 600          |            |              |                    |
| DB                    | 541  | EHMISVSAFTIMEGEPPTVLSVTRQOVPSIKIINYKNISSSILLYWDPPPEYNGKIT         | 600          |            |              |                    |
| QY                    | 601  | HYTIYAMELDTNRAFOITIDNSFLITGLKKTYYKORVAASTHDGESSLSBENDIFVRT        | 660          |            |              |                    |
| DB                    | 601  | HYTIYAMELDTNRAFOITIDNSFLITGLKKTYYKORVAASTHDGESSLSBENDIFVRT        | 660          |            |              |                    |
| QY                    | 661  | SEDEPESSQDVEVIDVDTADEIRLKWSPPEKNGIIIAEVLKYKNIDTLYMKNTSTTDII       | 720          |            |              |                    |
| DB                    | 661  | SEDEPESSQDVEVIDVDTADEIRLKWSPPEKNGIIIAEVLKYKNIDTLYMKNTSTTDII       | 720          |            |              |                    |
| QY                    | 721  | LRNLRPHLTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGEIELSF    | 780          |            |              |                    |
| DB                    | 721  | LRNLRPHLTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGEIELSF    | 780          |            |              |                    |
| QY                    | 781  | LPPSSPNGIIKKYTIYLRKSGNNEERTINTSLTQNIKVKYQYIIIEVSASTLKGEV          | 840          |            |              |                    |
| DB                    | 781  | LPPSSPNGIIKKYTIYLRKSGNNEERTINTSLTQNIKVKYQYIIIEVSASTLKGEV          | 840          |            |              |                    |
| QY                    | 841  | RSAPISILTEEDAPSDPPQDPSVKQLSGVTVKLSWQPPLEPGCIILYTYTVYVWNRSSSLKT    | 900          |            |              |                    |
| DB                    | 841  | RSAPISILTEEDAPSDPPQDPSVKQLSGVTVKLSWQPPLEPGCIILYTYTVYVWNRSSSLKT    | 900          |            |              |                    |
| QY                    | 901  | INVTETSLSDLDYNNVEYSAYTASTRFGDGKTSNIIISFQTPEGAPSDPPKDVYANL         | 960          |            |              |                    |
| DB                    | 901  | INVTETSLSDLDYNNVEYSAYTASTRFGDGKTSNIIISFQTPEGAPSDPPKDVYANL         | 960          |            |              |                    |
| QY                    | 961  | SSSSIIILFWTPSPKNGIIQYYSVYVRNTSGTFMQNFTHLNTDNDNNMTVSTIIDKLTJ       | 1020         |            |              |                    |
| DB                    | 961  | SSSSIIILFWTPSPKNGIIQYYSVYVRNTSGTFMQNFTHLNTDNDNNMTVSTIIDKLTJ       | 1020         |            |              |                    |
| QY                    | 1021 | FSYTYFWLTASTSVGNKSSDIIEVYTDQDIPGEGVGNLTYESISSSTAINVSWPPAQ         | 1080         |            |              |                    |
| DB                    | 1021 | FSYTYFWLTASTSVGNKSSDIIEVYTDQDIPGEGVGNLTYESISSSTAINVSWPPAQ         | 1080         |            |              |                    |
| QY                    | 1081 | NGLYFYVSLILQOQPRHVRPPLVYERSIYFDNILEKYTDYILKITPSTKGFSDTYTAQ        | 1140         |            |              |                    |
| DB                    | 1081 | NGLYFYVSLILQOQPRHVRPPLVYERSIYFDNILEKYTDYILKITPSTKGFSDTYTAQ        | 1140         |            |              |                    |
| QY                    | 1141 | LYIKTEEDVPETSPILNTFFKNLSSTVLLSWDPVVKPKNGAIISYDOLTQGNENYSFITS      | 1200         |            |              |                    |
| DB                    | 1141 | LYIKTEEDVPETSPILNTFFKNLSSTVLLSWDPVVKPKNGAIISYDOLTQGNENYSFITS      | 1200         |            |              |                    |
| QY                    | 1201 | DNYIILBELSPPTLYSFFAAARTRKGLGFPSSILFPYTTDESVPPLAPQNLTLINCTSDFW     | 1260         |            |              |                    |
| DB                    | 1201 | DNYIILBELSPPTLYSFFAAARTRKGLGFPSSILFPYTTDESVPPLAPQNLTLINCTSDFW     | 1260         |            |              |                    |
| QY                    | 1261 | LKWSPSPLPGGIVKYVSKIHHEHTDTIYKNIISGFKTEAKLVLEPVSTYSIRVSAFTK        | 1320         |            |              |                    |
| DB                    | 1261 | LKWSPSPLPGGIVKYVSKIHHEHTDTIYKNIISGFKTEAKLVLEPVSTYSIRVSAFTK        | 1320         |            |              |                    |
| QY                    | 1321 | VGNQGNFSNVKFTTQESVDPVWQNMCMATSWQSVLVKWDPPKKANGIIIOYMYTVERN        | 1380         |            |              |                    |
| DB                    | 1321 | VGNQGNFSNVKFTTQESVDPVWQNMCMATSWQSVLVKWDPPKKANGIIIOYMYTVERN        | 1380         |            |              |                    |
| QY                    | 1381 | STKVSPOQDMYTFIKLLANTSVYFKVRASTASAGEDESTCHVSTILPETVPSVPTNIAFSD     | 1440         |            |              |                    |
| DB                    | 1381 | STKVSPOQDMYTFIKLLANTSVYFKVRASTASAGEDESTCHVSTILPETVPSVPTNIAFSD     | 1440         |            |              |                    |
| QY                    | 1441 | VOSTSATLWIRPDTILGYFQNYKITTLQRAQCKEWESECEVYQKIQYLYEAHLTBEET        | 1500         |            |              |                    |
| DB                    | 1441 | VOSTSATLWIRPDTILGYFQNYKITTLQRAQCKEWESECEVYQKIQYLYEAHLTBEET        | 1500         |            |              |                    |
| QY                    | 1501 | YVGLKFPWYRFOVAASTNAGYGNASNIWSTKTLPGPDGPPENVHVATSPFSISWS           | 1560         |            |              |                    |
| DB                    | 1501 | YVGLKFPWYRFOVAASTNAGYGNASNIWSTKTLPGPDGPPENVHVATSPFSISWS           | 1560         |            |              |                    |
| QY                    | 1561 | BEAVITGPTCYLIDVKSVDNDEFNISFKSNEENTKIEIKOLEIFTRYSVVITATFNGIS       | 1620         |            |              |                    |
| DB                    | 1561 | BEAVITGPTCYLIDVKSVDNDEFNISFKSNEENTKIEIKOLEIFTRYSVVITATFNGIS       | 1620         |            |              |                    |
| QY                    | 1621 | AAVTEGKSAEMIIVTTLESAPKDPNNMTFOKIPDEVTKFQTLFPLPPSQPNNGNIQVYQAL     | 1680         |            |              |                    |
| DB                    | 1621 | AAVTEGKSAEMIIVTTLESAPKDPNNMTFOKIPDEVTKFQTLFPLPPSQPNNGNIQVYQAL     | 1680         |            |              |                    |
| QY                    | 1681 | VYREDDPTAVQIHNLSIIQKNTTFVIAMLEGLKGGHTYNIWYVYNSAGAGKVPMTIM         | 1740         |            |              |                    |
| DB                    | 1681 | VYREDDPTAVQIHNLSIIQKNTTFVIAMLEGLKGGHTYNIWYVYNSAGAGKVPMTIM         | 1740         |            |              |                    |
| QY                    | 1741 | DIKAPARPKTKPTPIYDATGKLLVTSTTTIRMPICYSDDDHGPICKNVQVLATETGAQHD      | 1800         |            |              |                    |
| DB                    | 1741 | DIKAPARPKTKPTPIYDATGKLLVTSTTTIRMPICYSDDDHGPICKNVQVLATETGAQHD      | 1800         |            |              |                    |
| QY                    | 1801 | GNVTWYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYIIIGADNACMIPGNEDEKIC       | 1860         |            |              |                    |
| DB                    | 1801 | GNVTWYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYIIIGADNACMIPGNEDEKIC       | 1860         |            |              |                    |
| QY                    | 1861 | NGPLKPKKQYLFKPFATNIMGOTDSDYSDPVKTLGEGLSERTVEIILSVTLICILSILL       | 1920         |            |              |                    |
| DB                    | 1861 | NGPLKPKKQYLFKPFATNIMGOTDSDYSDPVKTLGEGLSERTVEIILSVTLICILSILL       | 1920         |            |              |                    |
| QY                    | 1921 | GTAIFAFARIROKQEGGYTSPQDAEIIDTKLKDQLITVADLELKDRLTR-----            | 1972         |            |              |                    |
| DB                    | 1921 | GTAIFAFARIROKQEGGYTSPQDAEIIDTKLKDQLITVADLELKDRLTR-----            | 1972         |            |              |                    |
| QY                    | 1973 | -PISKSKSFLQHVVEELCTNNLKFQEFSELPKFLODLSSTDADLPWNRAKNRFPNIKPN       | 2031         |            |              |                    |
| DB                    | 1981 | KFISKSKSFLQHVVEELCTNNLKFQEFSELPKFLODLSSTDADLPWNRAKNRFPNIKPN       | 2039         |            |              |                    |
| QY                    | 2032 | NNNRVKLTADASVPGSDYINASYISGYLCPNEFIATQGPLCTGVDFWVWVETRAKTIV        | 2091         |            |              |                    |
| DB                    | 2040 | NNNRVKLTADASVPGSDYINASYISGYLCPNEFIATQGPLCTGVDFWVWVETRAKTIV        | 2099         |            |              |                    |
| QY                    | 2092 | MLTQCFEKGRIRCHOYWPEDNKPVTVFGDIVITIKLMEDVQIDWTIRDLKIERHGDCTVR      | 2151         |            |              |                    |

Db 2100 MLTQCPEKGRIRCHQWPENDKPVTIVFGDIIIVITKLMEDVQIDWTIRDLKIERHGDCMTVR 2159  
Qy 2152 QCNFTAWPEHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGTGTFVIALDHLTQH 2211  
Db 2160 QCNFTAWPEHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGTGTFVIALDHLTQH 2219  
Qy 2212 INDHDFVDIYGLVAELRSRMCVMQNLQAYIFLHQICILLNKGNSQNPICFVNTSALOK 2271  
Db 2220 INDHDFVDIYGLVAELRSRMCVMQNLQAYIFLHQICILLNKGNSQNPICFVNTSALOK 2279  
Qy 2272 MDSLDAWEGDVELEWBEETM 2291  
Db 2280 MDSLDAWEGDVELEWBEETM 2299

## RESULT 3

AAE37320  
ID AAE37320 standard; protein; 2299 AA.

XX AAE37320;

XX 07-AUG-2003 (first entry)

DE Human receptor tyrosine kinase protein #1.

XX Human; receptor tyrosine phosphatase; diabetes; obesity; CNS disorder;  
KW multiple sclerosis; epilepsy; chronic obstructive pulmonary disease;  
KW COPD; Parkinson's disease; stroke; cardiovascular disorder; arrhythmia;  
KW congestive heart failure; myocardial infarction; ischaemic heart disease;  
KW gene therapy; anorectic; cardiatic; neuroprotective; anticonvulsant;  
KW cerebroprotective; vasotropic; antiarrhythmic; receptor; enzyme;  
KW chromosome 12.

XX Homo sapiens.

XX WO2003033688-A1.

PN 24-APR-2003.

PD 14-OCT-2002; 2002WO-BP011473.

PP 16-OCT-2001; 2001US-0329329P.

XX (PARB ) BAYER AG.

XX Liou J;

XX WPI; 2003-403215/38.

DR N-PSDB; AAD56411.

XX Novel polynucleotides encoding human receptor tyrosine phosphatase

PT polypeptides, useful for treating diabetes, CNS disorders, obesity,

PT chronic obstructive pulmonary disease and cardiovascular disorders.

XX Claim 1; Fig 2; 163pp; English.

CC The present invention relates to receptor tyrosine phosphatase proteins  
CC and polynucleotides encoding them. Sequences of the invention are useful  
CC in the preparation of medicaments for modulating the activity of receptor  
CC tyrosine phosphatase in disease such as diabetes, obesity, CNS disorders  
CC (multiple sclerosis, epilepsy, Parkinson's disease and stroke), chronic  
CC obstructive pulmonary disease (COPD) and cardiovascular disorders (e.g.  
CC congestive heart failure, myocardial infarction, ischaemic heart disease  
CC and arrhythmia). They are also used in gene therapy. The present sequence  
CC is human receptor tyrosine kinase protein. Receptor tyrosine kinase gene  
CC is located on chromosome 12

XX Sequence 2299 AA;

Query Match 99.3%; Score 11905; DB 6; Length 2299;

Best Local Similarity 99.2%; Pred. No. 0;

Matches 2282; Conservative 3; Mismatches 5; Indels 10; Gaps 2;

Qy 1 MDPLIIFLLFTGTSETQVDSNVVPGTRYDITISSISTYTSPTVTRIVTPNVTKPGPV 60  
Db 1 MDPLIIFLLFTGTSETQVDSNVVPGTRYDITISSISTYTSPTVTRIVTPNVTKPGPV 60  
Qy 61 FLAGERVCSAGILLSSWNTPPNPGRIISYIVKYKEVCPMMQVTVTVTSKPSDSLEVLN 120  
Db 61 FLAGERVCSAGILLSSWNTPPNPGRIISYIVKYKEVCPMMQVTVTVTSKPSDSLEVLN 120  
Qy 121 LNPGETTYEIKVAENSAGIGVFSDFLFTQTAESAFAKVVNLTVZAYNASAVKLIWYLP 180  
Db 121 LNPGETTYEIKVAENSAGIGVFSDFLFTQTAESAFAKVVNLTVZAYNASAVKLIWYLP 180  
Qy 181 PNGKITSPKISVKHARSGIVKDVSIKVEDILTKLPESNENSESFLMSTASPTLGRV 240  
Db 181 PNGKITSPKISVKHARSGIVKDVSIKVEDILTKLPESNENSESFLMSTASPTLGRV 240  
Qy 241 TPPSETHTSSSTLTQNEISSWKEPISFVVTTHLRPYTYLFEVSAATTEAGYIDSTIV 300  
Db 241 TPPSETHTSSSTLTQNEISSWKEPISFVVTTHLRPYTYLFEVSAATTEAGYIDSTIV 300  
Qy 301 PESVPEGPQNCVTGNTGKSFILWDPPITVTKFSYRVELYGPGRILDNSTDKLKA 360  
Db 301 PESVPEGPQNCVTGNTGKSFILWDPPITVTKFSYRVELYGPGRILDNSTDKLKA 360  
Qy 361 FTNLTPFTMYDYIAAETSAGTGPKNISVFTPPDVGAVFDLQLAEVSTQVRIWTKP 420  
Db 361 FTNLTPFTMYDYIAAETSAGTGPKNISVFTPPDVGAVFDLQLAEVSTQVRIWTKP 420  
Qy 421 ROPNGIINQYRVKVLVPETGILENTLTGNNEYINDMPAIEINIVEMVGLYEGSASM 480  
Db 421 ROPNGIINQYRVKVLVPETGILENTLTGNNEYINDMPAIEINIVEMVGLYEGSASM 480  
Qy 481 SSDLSLATFIYNHSHDPKFNPAENAEQTSVVTTRNQYITDIAAEQLSYVIRLVPT 540  
Db 481 SSDLSLATFIYNHSHDPKFNPAENAEQTSVVTTRNQYITDIAAEQLSYVIRLVPT 540  
Qy 541 EHMISVSAFTIMGEPTVLSVTRTQVPPSSIKIINYKNISSSILLYWDPEYNGKIT 600  
Db 541 EHMISVSAFTIMGEPTVLSVTRTQVPPSSIKIINYKNISSSILLYWDPEYNGKIT 600  
Qy 601 HYTIYAMELDNRAFOITIDNSFLITGLKYYTKYKMRVAATHDGESSLSENDIFV 660  
Db 601 HYTIYAMELDNRAFOITIDNSFLITGLKYYTKYKMRVAATHDGESSLSENDIFV 660  
Qy 661 SEDPESSQDVEVDVDAEIRLKWSPPEKNGIIIAEVLYKNDITLYMKNSTTDDII 720  
Db 661 SEDPESSQDVEVDVDAEIRLKWSPPEKNGIIIAEVLYKNDITLYMKNSTTDDII 720  
Qy 721 LRNLPHLTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSGIELSF 780  
Db 721 LRNLPHLTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSGIELSF 780  
Qy 781 LPPSSPNGIIKKYTIYLRSGNBERNTINTSLTONIKVLYKYTOYIIEVSASTLKGEV 840  
Db 781 LPPSSPNGIIKKYTIYLRSGNBERNTINTSLTONIKVLYKYTOYIIEVSASTLKGEV 840  
Qy 841 RSAPISILTEEDAPDPSQDFSVKOLSGVTKLSWQPLEPNGIILYTVVYVWNSSLKT 900  
Db 841 RSAPISILTEEDAPDPSQDFSVKOLSGVTKLSWQPLEPNGIILYTVVYVWNSSLKT 900  
Qy 901 INVETSLSDLDYNNVYSAYVTAFTFGDKTGSNIISFOTPEGADSDPPKDVYANL 960  
Db 901 INVETSLSDLDYNNVYSAYVTAFTFGDKTGSNIISFOTPEGADSDPPKDVYANL 960  
Qy 961 SSSSIIILFWTPPSKPNNGIIQYYSVYRNTSGTFMQNFTLHETLNDPDMVTSTIIDK 1020  
Db 961 SSSSIIILFWTPPSKPNNGIIQYYSVYRNTSGTFMQNFTLHETLNDPDMVTSTIIDK 1020  
Qy 1021 PSYTYFWLTASTSVGNKNSDIIIEVYTDQDIPGFCVGLTYESISSSTAINVSWVPPAQ 1080  
Db 1021 PSYTYFWLTASTSVGNKNSDIIIEVYTDQDIPGFCVGLTYESISSSTAINVSWVPPAQ 1080

QY 1081 NGLVFFVYVSLILQOTPRHVRPPLVTVYERSIYFDNLEKYTDYILKITPSTKGFSDTYTAQ 1140  
DB 1081 NGLVFFVYVSLILQOTPRHVRPPLVTVYERSIYFDNLEKYTDYILKITPSTKGFSDTYTAQ 1140  
QY 1141 LYIKTEEDVPETSPINTPKNSSTSVLLSWPPVPKNGAIIISYDLTLQGNENYSFITS 1200  
DB 1141 LYIKTEEDVPETSPINTPKNSSTSVLLSWPPVPKNGAIIISYDLTLQGNENYSFITS 1200  
QY 1201 DNYIILEELSPFTLYSFFFAAARKGLGPSSILFFYTDSEVPLAPPQNTLINCTSDFVM 1260  
DB 1201 DNYIILEELSPFTLYSFFFAAARKGLGPSSILFFYTDSEVPLAPPQNTLINCTSDFVM 1260  
QY 1261 LKWSPSPLPGGIVKVSFKIHEHETDTIYKNIISGFKTEAKLVGLEPVSYSIRVSAFTK 1320  
DB 1261 LKWSPSPLPGGIVKVSFKIHEHETDTIYKNIISGFKTEAKLVGLEPVSYSIRVSAFTK 1320  
QY 1321 VGNQOFSNVKFTTQESVDPVQNNQCMATSNQSVLVKWDPPKANGIITQYMTVERN 1380  
DB 1321 VGNQOFSNVKFTTQESVDPVQNNQCMATSNQSVLVKWDPPKANGIITQYMTVERN 1380  
QY 1381 STKVSQDHYMTFIKLLANTSYVFKVRASTSAEGDDESTCHVSTLPETVPSVPTIAFSD 1440  
DB 1381 STKVSQDHYMTFIKLLANTSYVFKVRASTSAEGDDESTCHVSTLPETVPSVPTIAFSD 1440  
QY 1441 VOSTSATLTWIRPDTILGFQNYKIITQLRAQCKEWESEECVEYQKIQYLYEAHLTEST 1500  
DB 1441 VOSTSATLTWIRPDTILGFQNYKIITQLRAQCKEWESEECVEYQKIQYLYEAHLTEST 1500  
QY 1501 VYGLKKFRWYRFOVAASSTNAGYGNASNWIISTKTLPGPPDGPENVHVATSPFSISWS 1560  
DB 1501 VYGLKKFRWYRFOVAASSTNAGYGNASNWIISTKTLPGPPDGPENVHVATSPFSISWS 1560  
QY 1561 EPAVITGPTCYLIDVKSVDNDENISFIKSNEENKTIEIKOLEIFTRYSVVITAFNGNIS 1620  
DB 1561 EPAVITGPTCYLIDVKSVDNDENISFIKSNEENKTIEIKOLEIFTRYSVVITAFNGNIS 1620  
QY 1621 AAYVEGKSSAEMIVTTLESAPDPPNNMTFKQIPDEVTKFQLTFLPPSPNGNIQYQAL 1680  
DB 1621 AAYVEGKSSAEMIVTTLESAPDPPNNMTFKQIPDEVTKFQLTFLPPSPNGNIQYQAL 1680  
QY 1681 VYREDDPTAVQIHNLSIIQKNTFTVIAMLEGLKGGHTYNSVYVNSAGAGKVPWRITM 1740  
DB 1681 VYREDDPTAVQIHNLSIIQKNTFTVIAMLEGLKGGHTYNSVYVNSAGAGKVPWRITM 1740  
QY 1741 DIKAPARPKTPPIYDATGKLLVSTTTIRMPICYSDDHGPIKNQVVLATETGAQHD 1800  
DB 1741 DIKAPARPKTPPIYDATGKLLVSTTTIRMPICYSDDHGPIKNQVVLATETGAQHD 1800  
QY 1801 GNVTKWYDAYFNKARPYFTNEGPPNPPCTEGTKFSGNEEIIYIIGADNACMIPGNEDKIC 1860  
DB 1801 GNVTKWYDAYFNKARPYFTNEGPPNPPCTEGTKFSGNEEIIYIIGADNACMIPGNEDKIC 1860  
QY 1861 NGPLKPKQYLFKFRATNMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCISILL 1920  
DB 1861 NGPLKPKQYLFKFRATNMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCISILL 1920  
QY 1921 GTAFAPAFARIQKQEGGTYSPQDAEIIIDTKLKDQILTVADLELKDRLTR----- 1972  
DB 1921 GTAFAPAFARIQKQEGGTYSPQDAEIIIDTKLKDQILTVADLELKDRLTR----- 1972  
QY 1973 -PTSKSFLQHVBEELCTNNLKFQEFSELPLFKPLQDLSSTDADLPWNRKRNFPNPKPN 2031  
DB 1981 KPISKKSFLQHVBEELCTNNLKFQEFSELPLFKPLQDLSSTDADLPWNRKRNFPNPKPN 2039  
QY 2032 NNNRVKLIADASVPGSDYINASYISGLCPNEFIATQGPLPGTVGDFWPMVWETRAKTIV 2091  
DB 2040 NNNRVKLIADASVPGSDYINASYISGLCPNEFIATQGPLPGTVGDFWPMVWETRAKTIV 2099  
QY 2092 MLTQCCEKGRIRCHOYWPEDNKVTVFGDIVITKLMEVDQIDWITRDLEKIERHGDQTVR 2151  
DB 2100 MLTQCCEKGRIRCHOYWPEDNKVTVFGDIVITKLMEVDQIDWITRDLEKIERHGDQTVR 2159  
QY 2152 QCNFTAWPEHGVPENSAPLIHFVKLVRASRAHDTTPMIVHCSAGVGRGTGFIALDHLTQH 2211

DB 2160 QCNFTAWPEHGVPENSAPLIHFVKLVRASRAHDTTPMIVHCSAGVGRGTGFIALDHLTQH 2219  
QY 2212 INDHDFVDIYGLVABLSRERMCMQNLAQYIFLHCILDLISNKGSNOPICFVNVYSALQK 2271  
DB 2220 INDHDFVDIYGLVABLSRERMCMQNLAQYIFLHCILDLISNKGSNOPICFVNVYSALQK 2279  
QY 2272 MDSLAMEGDGVELEWSETTM 2291  
DB 2280 MDSLAMEGDGVELEWSETTM 2299  
RESULT 4  
AAO18738  
ID AAO18738 standard; protein; 2300 AA.  
XX  
AC AAO18738;  
XX  
DT 24-OCT-2002 (first entry)  
XX Human NOV2c protein.  
DE Human; NOVX; autoimmune disease; cancer; infection; inflammatory disease;  
KW storage disorder; muscle disorder; neurodegenerative disorder; nootropic;  
KW developmental defect; neuroprotective; antiparkinsonian; hypotensive;  
KW hypertensive; haemostatic; cardiast; antidiabetic; dermatological;  
KW immunosuppressive; antiinflammatory; virucide; antibacterial; anti-HIV;  
KW antiparasitic; antiallergic; antiasthmatic; antirheumatic; antiarthritic;  
KW vulnary; anorectic; antidiabetic; immunomodulator; antipsoriatic;  
KW nephrotropic; kerolytic; antiulcer; cerabroprotective; anticonvulsant;  
KW antifertility; antimanic; antidepressant; metabolic; cytostatic;  
KW tranquilizer; analgesic.  
XX Homo sapiens.  
OS  
XX WO200257450-A2.  
XX 25-JUL-2002.  
XX 29-NOV-2001; 2001WO-US048922.  
XX 29-NOV-2000; 2000US-0253834P.  
PR 30-NOV-2000; 2000US-0250926P.  
PR 25-JAN-2001; 2001US-0264180P.  
PR 20-AUG-2001; 2001US-0313656P.  
PR 05-OCT-2001; 2001US-0327456P.  
PR 28-NOV-2001; 2001US-00327456.  
XX (CURA-) CURAGEN CORP.  
PA  
XX Edinger S, Macdougall JR, Millet I, Ellerman K, Stone DJ;  
PI Gerlach V, Grosse WM, Alsobrook JP, Lepley DM, Rieger D, Burgeas CE;  
PI Casman SJ, Spytek KA, Boldog FL, Li L, Padigar M, Mishra V;  
PI Patturajan M, Shenoy S, Rastelli L, Tchernev VT, Vernet CAM;  
PI Zerhusen BD, Malyankar UM, Guo X, Miller CE, Gangolli EA;  
XX WPI: 2002-590741/63.  
DR N-PSDB; ABO6283.  
XX  
PT Novel isolated polypeptide, designated NOVX, useful for treating or  
preventing in NOVX-associated disorders e.g. cardiomyopathy,  
atherosclerosis, diabetes, cancer, allergy, asthma, Crohn's disease.  
PT  
XX Claim 1; Page 30; 353pp; English.  
PS  
XX The present invention provides the protein and coding sequences of  
several novel human proteins, designated NOVX. These can be used in the  
treatment of, amongst others, cancers, autoimmune diseases, infections,  
CC inflammatory diseases, storage disorders, muscle disorders,  
CC neurodegenerative diseases and developmental defects. The present  
sequence is a protein of the invention  
XX  
SQ Sequence 2300 AA;

|                       |                  |  |            |              |  |
|-----------------------|------------------|--|------------|--------------|--|
| Query Match           | 98.5%;           | Score 11807.5;   | DB 5;      | Length 2300; |  |
| Best local Similarity | 98.7%;           | Pred. No. 0;   |            |              |  |
| Matches 2276;         | Conservative .5; | Mismatches 5;  | Indels 19; | Gaps 9;      |  |
| QY                    | 1                | MDFLIFLLFLLFICTSETQVDVSNVPGTRDYDITISSITSTYTSPVTRIVTNPVTKGPPV   | 60         |              |  |
| DB                    | 1                | MDFLIFLLFLLFICTSETQVDVSNVPGTRDYDITISSITSTYTSPVTRIVTNPVTKGPPV   | 60         |              |  |
| QY                    | 61               | FLAGERVGSAGILLSNWTPNPNNGRIISYIVKYKEVCPMMQTYTVQVRSKPDLEVLTLN    | 120        |              |  |
| DB                    | 61               | FLAGERVGSAGILLSNWTPNPNNGRIISYIVKYKEVCPMMQTYTVQVRSKPDLEVLTLN    | 120        |              |  |
| QY                    | 121              | LNPGTTYEIKVAENAGAGVFDPLFQTAES-APGKVNLTVAYNANSAVKLIWYLR         | 179        |              |  |
| DB                    | 121              | LNPGTTYEIKVAENAGAGVFDPLFQTAES-APGKVNLTVAYNANSAVKLIWYLR         | 180        |              |  |
| QY                    | 180              | QPNKGTSPKISVKHARSGLVVDKSVRVEDILTGLPECN-RNSSEFLWSTASPSPTLG      | 238        |              |  |
| DB                    | 181              | QPNKGTSPKISVKHARSGLVVDKSVRVEDILTGLPECN-RNSSEFLWSTASPSPTLG      | 240        |              |  |
| QY                    | 239              | RVTTPSRTHSSSTLTQNEISSVWKEPIFPVVTTLRPYTYTLFEVSAATTEAGYIDSTIV    | 298        |              |  |
| DB                    | 241              | RVTTPSRTHSSSTLTQNEISSVWKEPIFPVVTTLRPYTYTLFEVSAATTEAGYIDSTIV    | 300        |              |  |
| QY                    | 299              | RTPEVPEGPONCVTGNITGKSFSLWDPPPTIVTGKFSYRVLYGPGSRILDNSTKDLK      | 358        |              |  |
| DB                    | 301              | RTPEVPEGPONCVTGNITGKSFSLWDPPPTIVTGKFSYRVLYGPGSRILDNSTKDLK      | 360        |              |  |
| QY                    | 359              | FAPTNLTPTTMDVYIAAETSAGTGPKSNISVTPDPVPGAVFDLQAEVESTQVRIWK       | 418        |              |  |
| DB                    | 361              | FAPTNLTPTTMDVYIAAETSAGTGPKSNISVTPDPVPGAVFDLQAEVESTQVRIWK       | 420        |              |  |
| QY                    | 419              | KRPQNGIINOXKVKVLVPETGIIILENTLLTGNEVINDPMAPEIINIVEPMWGLYEGSA    | 478        |              |  |
| DB                    | 421              | KRPQNGIINOXKVKVLVPETGIIILENTLLTGNE-INDPMAPEIINIVEPMWGLYEGSA    | 479        |              |  |
| QY                    | 479              | EMSSDLHSLATFIYNHSHDPKPNPARNRAEDQTSPPVVTTRNOYITDIAAEQLSVIRRLVP  | 538        |              |  |
| DB                    | 480              | EMSSDLHSLATFIYNHSHDPKPNPARNRAEDQTSPPVVTTRNOYITDIAAEQLSVIRRLVP  | 539        |              |  |
| QY                    | 539              | FTEHMTSVSAGFTMGSGPPTVLSVRTRQOVPSIKIINYNKISSSSILLYWDPPEYNGK     | 598        |              |  |
| DB                    | 540              | FTEHMTSVSAGFTMGSGPPTVLSVRTRQOVPSIKIINYNKISSSSILLYWDPPEYNGK     | 599        |              |  |
| QY                    | 599              | ITHYTIYAMELDTNRAFOITTDINSFLIT--GLKYYTKYKMRVAASTHDSGSSLSSENDI   | 656        |              |  |
| DB                    | 600              | ITHYTIYAMELDTNRAFOITTDINSFLIT--GLKYYTKYKMRVAASTHDSGSSLSSENDI   | 659        |              |  |
| QY                    | 657              | FVRTSEDESSPDQVEVIDVTADERLKWSPPEKPNNGIIIAVEVLYKNIDTLYMKNTST     | 716        |              |  |
| DB                    | 660              | FVRTSEDESSPDQVEVIDVTADERLKWSPPEKPNNGIIIAVEVLYKNIDTLYMKNTST     | 719        |              |  |
| QY                    | 717              | TDIILNLRPHLTLYNLSVRSYTRFPGHGNQVSSLLSVRTSETVPSAPENITYKNISSGEI   | 776        |              |  |
| DB                    | 720              | TDIILNLRPHLTLYNLSVRSYTRFPGHGNQVSSLLSVRTSETVPSAPENITYKNISSGEI   | 779        |              |  |
| QY                    | 777              | ELSFLEPPSPNGIIGKYYTLYLKEBSNGNEERTINTTSLTONIKVKYTOYIIEVSASTLK   | 836        |              |  |
| DB                    | 780              | ELSFLEPPSPNGIIGKYYTLYLKEBSNGNEERTINTTSLTONI--LKYYTOYIIEVSASTLK | 837        |              |  |
| QY                    | 837              | GEGRVSAPTSILTEEDAPSPDPQFSVKOLSGVTVKLSWQPPLEPNNGIILYTVYVW-NR    | 895        |              |  |
| DB                    | 838              | GEGRVSAPTSILTEEDAPSPDPQFSVKOLSGVTVKLSWQPPLEPNNGIILYTVYVWNR     | 897        |              |  |
| QY                    | 896              | SSLKTTINVTETSLSDLDYNVEYSAYVTASTRFGDGKTCGNSIISFQTPPEGAPSPDKDV   | 955        |              |  |
| DB                    | 898              | SSLKTTINVTETSLSDLDYNVEYSAYVTASTRFGDGKTCGNSIISFQTPPEG-PSDPPKDV  | 956        |              |  |
| QY                    | 956              | YANLSSSSIIILFWTPPPKNGIIOYYSVYRNTSGTFMQNFTLHETDNDFDMVTSTII      | 1015       |              |  |
| DB                    | 957              | YANLSSSSIIILFWTPPPKNGIIOYYSVYRNTSGTFMQNFTLHETDNDFDMVTSTII      | 1016       |              |  |

|    |      |  |      |  |  |
|----|------|--|------|--|--|
| QY | 1016 | DKLTIFSYTTFWLTASTSVGNGKSSDIIIEVTTDQDIPGPFVGNLTYESISSTAINVSW    | 1075 |  |  |
| DB | 1017 | DKLTIFSYTTFWLTASTSVGNGKSSDIIIEVTTDQDIPGPFVGNLTYESISSTAINVSW    | 1076 |  |  |
| QY | 1076 | PPAQNGLVYVYVSLILOOTPRHVRPPLVYERSYFVFNLEKVTYDILKITPSTEGFSD      | 1135 |  |  |
| DB | 1077 | PPAQNGLVYVYVSLILOOTPRHVRPPLVYERSYFVFNLEKVTYDILKITPSTEGFSD      | 1136 |  |  |
| QY | 1136 | TYTAQLYIKTEEDVPETSPINTPKNLSSTSVLLSWDPVPKNGALISYDLTLQGNENY      | 1195 |  |  |
| DB | 1137 | TYTAQLYIKTEEDVPETSPINTPKNLSSTSVLLSWDPVPKNGALISYDLTLQGNENY      | 1196 |  |  |
| QY | 1196 | SPITSDNYIILBELSPFTLYSPFAAARTKGLGPSSILFFYTDSEVPLAPPQNLTLINCT    | 1255 |  |  |
| DB | 1197 | SPITSDNYIILBELSPFTLYSPFAAARTKGLGPSSILFFYTDSEVPLAPPQNLTLINCT    | 1256 |  |  |
| QY | 1256 | SPFVWLKWSPLPGIIVKVYFKIHEHETDTIYKNIISGFKTEAKLVGLEPVSTYSIRV      | 1315 |  |  |
| DB | 1257 | SPFVWLKWSPLPGIIVKVYFKIHEHETDTIYKNIISGFKTEAKLVGLEPVSTYSIRV      | 1316 |  |  |
| QY | 1316 | SAFTKVGNGNQFSNVVKFTTQESVDPVQNMOCMATSWQSVLVKMDPPKKANGIITQYMW    | 1375 |  |  |
| DB | 1317 | SAFTKVGNGNQFSNVVKFTTQESVDPVQNMOCMATSWQSVLVKMDPPKKANGIITQYMW    | 1376 |  |  |
| QY | 1376 | TVERNSTKVSPODHMYTFIKLLANTSYVPKVAASTSAGEGDESTCHVSTLPETVPVPTN    | 1435 |  |  |
| DB | 1377 | TVERNSTKVSPODHMYTFIKLLANTSYVPKVAASTSAGEGDESTCHVSTLPETVPVPTN    | 1436 |  |  |
| QY | 1436 | IAPSDVQSTSATLTWIRPDTILGYFQNYKIITTLQRAQCKEWESECEVQKIQLYLEAH     | 1495 |  |  |
| DB | 1437 | IAPSDVQSTSATLTWIRPDTILGYFQNYKIITTLQRAQCKEWESECEVQKIQLYLEAH     | 1496 |  |  |
| QY | 1496 | LTEETVYGLKFRWTRFVAASTNAGYGNASNIWSTKTLPGPPDGPENHVHVVATSPFSI     | 1555 |  |  |
| DB | 1497 | LTEETVYGLKFRWTRFVAASTNAGYGNASNIWSTKTLPGPPDGPENHVHVVATSPFSI     | 1556 |  |  |
| QY | 1556 | SISWSEPAVITGPTCYLIDVKSVDNDEFNISFIKNEENKTIIEIKOLEIFTRYSVVITAF   | 1615 |  |  |
| DB | 1557 | SISWSEPAVITGPTCYLIDVKSVDNDEFNISFIKNEENKTIIEIKOLEIFTRYSVVITAF   | 1616 |  |  |
| QY | 1616 | TGNIISAAVYEGKSSAEMIVTTLESAPKOPNNMTFKIPEVTYKQLTLPSPQPNNGIQ      | 1675 |  |  |
| DB | 1617 | TGNIISAAVYEGKSSAEMIVTTLESAPKOPNNMTFKIPEVTYKQLTLPSPQPNNGIQ      | 1676 |  |  |
| QY | 1676 | VYQALVYREDDPTAVQIHNLSIIQKNTNFIAMLEGLKGHTYNTSVYAVNSAGAGPKVP     | 1735 |  |  |
| DB | 1677 | VYQALVYREDDPTAVQIHNLSIIQKNTNFIAMLEGLKGHTYNTSVYAVNSAGAGPKVP     | 1736 |  |  |
| QY | 1736 | MRITWDIKAPARPKTPIYDATGKLLVTSTTIIRMPICYSDHGPKNVQVLATET          | 1795 |  |  |
| DB | 1737 | MRITWDIKAPARPKTPIYDATGKLLVTSTTIIRMPICYSDHGPKNVQVLATET          | 1796 |  |  |
| QY | 1796 | GAQHDGNTWKYDAYFNKARPYFTNEGPNPCTGKTKFSGNEEIIYIIGADNACMIPGN      | 1855 |  |  |
| DB | 1797 | GAQHDGNTWKYDAYFNKARPYFTNEGPNPCTGKTKFSGNEEIIYIIGADNACMIPGN      | 1856 |  |  |
| QY | 1856 | EDKICNGPLKPKQKQYLFKFRATNIMGFTDSYSDPVKTLGEGLSERTVEIILSVTLCL     | 1915 |  |  |
| DB | 1857 | EDKICNGPLKPKQKQYLFKFRATNIMGFTDSYSDPVKTLGEGLSERTVEIILSVTLCL     | 1916 |  |  |
| QY | 1916 | SIIILGTAIFAPARIRQKQGGTYSPODAEIIDTKLKDQITVADLEKDERLTR---        | 1972 |  |  |
| DB | 1917 | SIIILGTAIFAPARIRQKQGGTYSPODAEIIDTKLKDQITVADLEKDERLTR---        | 1976 |  |  |
| QY | 1973 | -----PISKSKSFLQHVBEELCTNNNLKFOEBSSELPKFLQDLSSTDAOLPMNRAKNRFPN  | 2026 |  |  |
| DB | 1977 | YKSKIKPISKSKSFLQHVBEELCTNNNLKFOEBSSELPKFLQDLSSTDAOLPMNRAKNRFPN | 2036 |  |  |
| QY | 2027 | IKPYNNNRVKLIADASVPGSDYNASYSGLVCNEFIATQGLPGTGVDFRWWVWETR        | 2086 |  |  |
| DB | 2037 | IKPY--NNNRVKLIADASVPGSDYNASYSGLVCNEFIATQGLPGTGVDFRWWVWETR      | 2095 |  |  |
| QY | 2087 | AKTLVMLTQCFEKGIRHQYWPEDNKPVTVPDGVITVKLMDVQIDMTIRDLKIERHGD      | 2146 |  |  |

```
Db      2096 AKTLVWLTCFEKGRIRCHQYWPEDNKPVTVGDIITVKLMEDVQIDWTIRDLKIERHGD 2155
Qy      2147 CMTVRCQCNFTAPPEHGVPENSAPLIHFVKLVRSRAHDTTPMIVHCSAGVGRGTGVFIALD 2206
Db      2156 CMTVRCQCNFTAPPEHGVPENSAPLIHFVKLVRSRAHDTTPMIVHCSAGVGRGTGVFIALD 2215
Qy      2207 HLTOHNDHDFVDIYGLVAELSRMCMVQNLQAQYIFLHQICLDLLSNKGSNQPICFVNY 2266
Db      2216 HLTOHNDHDFVDIYGLVAELSRMCMVQNLQAQYIFLHQICLDLLSNKGSNQPICFVNY 2275
Qy      2267 SALQKQMSLDAMEGDVELEWEETM 2291
Db      2276 SALQKQMSLDAMEGDVELEWEETM 2300

RESULT 5
ADH41625
ID      ADH41625 standard; protein; 2300 AA.
XX
AC      ADH41625;
XX
DT      25-MAR-2004 (first entry)
XX
DE      Novel human protein NOV15g.
XX
KW      cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;
KW      antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;
KW      antiparkinsonian; antiasthmatic; antiinfertility; cardiomyopathy;
KW      atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;
KW      multiple sclerosis; graft-versus-host disease; Alzheimer's disease;
KW      Parkinson's disease; asthma; fertility disorder; chromosome mapping;
KW      tissue typing; preventive medicine; pharmacogenomic; vaccine.
XX
OS      Homo sapiens.
XX
PN      WO2003102159-A2.
XX
XX      11-DEC-2003.
XX
PF      04-JUN-2003; 2003WO-US017573.
XX
PR      04-JUN-2002; 2002US-0385490P.
PR      04-JUN-2002; 2002US-0385615P.
PR      04-JUN-2002; 2002US-0385755P.
PR      05-JUN-2002; 2002US-0386041P.
PR      06-JUN-2002; 2002US-0386355P.
PR      06-JUN-2002; 2002US-0386357P.
PR      06-JUN-2002; 2002US-0386447P.
PR      06-JUN-2002; 2002US-0386459P.
PR      06-JUN-2002; 2002US-0386465P.
PR      06-JUN-2002; 2002US-0386864P.
PR      07-JUN-2002; 2002US-0386701P.
PR      07-JUN-2002; 2002US-0386796P.
PR      07-JUN-2002; 2002US-0386931P.
PR      07-JUN-2002; 2002US-0387078P.
PR      07-JUN-2002; 2002US-0387081P.
PR      07-JUN-2002; 2002US-0387083P.
PR      10-JUN-2002; 2002US-0387429P.
PR      10-JUN-2002; 2002US-0387540P.
PR      10-JUN-2002; 2002US-0387866P.
PR      11-JUN-2002; 2002US-0387606P.
PR      11-JUN-2002; 2002US-0387610P.
PR      11-JUN-2002; 2002US-0387659P.
PR      11-JUN-2002; 2002US-0387668P.
PR      11-JUN-2002; 2002US-0387696P.
PR      11-JUN-2002; 2002US-0387859P.
PR      12-JUN-2002; 2002US-0387934P.
PR      12-JUN-2002; 2002US-0387960P.
PR      12-JUN-2002; 2002US-0388022P.
PR      12-JUN-2002; 2002US-0388096P.
PR      12-JUN-2002; 2002US-0388432P.
PR      12-JUN-2002; 2002US-0388479P.
```

```
PR      13-JUN-2002; 2002US-0389123P.
PR      14-JUN-2002; 2002US-0389120P.
PR      14-JUN-2002; 2002US-0389146P.
PR      17-JUN-2002; 2002US-0389742P.
PR      18-JUN-2002; 2002US-0389604P.
PR      18-JUN-2002; 2002US-0389884P.
PR      19-JUN-2002; 2002US-0390006P.
PR      19-JUN-2002; 2002US-0390144P.
PR      19-JUN-2002; 2002US-0390209P.
PR      25-JUN-2002; 2002US-0391728P.
PR      06-AUG-2002; 2002US-0401628P.
PR      09-AUG-2002; 2002US-0402268P.
PR      12-AUG-2002; 2002US-0402822P.
PR      13-AUG-2002; 2002US-0403458P.
PR      15-AUG-2002; 2002US-0403617P.
PR      15-AUG-2002; 2002US-0403732P.
PR      26-AUG-2002; 2002US-0406182P.
PR      12-SEP-2002; 2002US-0410085P.
PR      13-SEP-2002; 2002US-0410505P.
PR      23-SEP-2002; 2002US-0412955P.
PR      30-SEP-2002; 2002US-0415195P.
PR      23-OCT-2002; 2002US-0420627P.
PR      23-OCT-2002; 2002US-0420718P.
PR      24-OCT-2002; 2002US-0420852P.
PR      31-OCT-2002; 2002US-0422750P.
PR      01-NOV-2002; 2002US-0423095P.
PR      05-NOV-2002; 2002US-0423748P.
XX
XX      (CURA-) CURAGEN CORP.
XX
PA      Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;
PI      Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;
PI      Ellerman K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;
PI      Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JU, Ji W, Kekuda R;
PI      Khrantsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;
PI      MacIachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;
PI      Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;
PI      Rieger DK, Rothenberg ME, Shenoy SG, Shinkets RA, Smithson G;
PI      Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;
PI      Wolenc AR, Zhong M, Zhong H;
XX
XX      WPI; 2004-053467/05.
XX      N-PSDB; ADH41624.
XX
XX      New NOVX polypeptides and nucleic acid molecules useful for preventing or
XX      treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
XX      atherosclerosis or diabetes, in chromosome mapping, tissue typing or in
XX      pharmacogenomics.
XX
XX      'Claim 2; SEQ ID NO 178; 1503pp; English.
XX
XX      The invention relates to 566 new isolated human polypeptides and their
XX      encoding genes, sequences that are at least 95% identical to these or
XX      sequences comprising one or more conservative substitutions in these. The
XX      polypeptide, polynucleotide and antibodies against the polypeptides are
XX      useful in diagnosing, treating or preventing NOVX-associated disorders,
XX      e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,
XX      diabetes, AIDS, multiple sclerosis, graft-versus-host disease,
XX      Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.
XX      The nucleic acids are further used as hybridization probes, in chromosome
XX      mapping, tissue typing, preventive medicine, and pharmacogenomics. The
XX      polypeptides are also useful as vaccines. This sequence represents an
XX      example of the polypeptide of the invention.
XX
XX      Sequence 2300 AA;
SQ
Query Match      98.5%; Score 11807.5; DB 8; Length 2300;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 2276; Conservative 5; Mismatches 5; Indels 19; Gaps 9;
Qy      1 MDFLIFLLFFGTSETQVDVSNVPGTRDYDITISSITTYTSPVTRIVTNVTFPGPV 60
Db      1 MDFLIFLLFFGTSETQVDVSNVPGTRDYDITISSITTYTSPVTRIVTNVTFPGPV 60
```

QY 61 FLAGERVGSAGILLSWTPNPNNGRIISYIVKKEVCWPMQVYTOVRSPKDSLEVLLTN 120  
DB 61 FLAGERVGSAGILLSWTPNPNNGRIISYIVKKEVCWPMQVYTOVRSPKDSLEVLLTN 120  
QY 121 LNPCTTYEIKVAABENSAGIVFSDPPLFQTAES-APGKVNLVTEAYNASAVKLIWYLP 179  
DB 121 LNPCTTYEIKVAABENSAGIVFSDPPLFQTAESAPGKVNLVTEAYNASAVKLIWYLP 180  
QY 180 QPKGKITSFKISVKHARSIGVVDVSRVEDILTKLPECN-ENSESFLMSTASPTLG 238  
DB 181 QPKGKITSFKISVKHARSIGVVDVSRVEDILTKLPECNVENSSESFLMSTASPTLG 240  
QY 239 RVTPPRTTHSSSTLQNEISSYWKBPISFWVTHLRPYTYILFEVGAATTEAGYIDSTIV 298  
DB 241 RVTPPRTTHSSSTLQNEISSYWKBPISFWVTHLRPYTYILFEVGAATTEAGYIDSTIV 300  
QY 299 RTPESVPEGPQNCVTGNIIGKSFSLMDPPTIVTGKFSRVLYGSPGRILNDSKDLK 358  
DB 301 RTPESVPEGPQNCVTGNIIGKSFSLMDPPTIVTGKFSRVLYGSPGRILNDSKDLK 360  
QY 359 FAFNLTPFTMYDYIAAETSAGTGPKSNISVFTPPDVGAVFDLQLAEVESQVRAITWK 418  
DB 361 FAFNLTPFTMYDYIAAETSAGTGPKSNISVFTPPDVGAVFDLQLAEVESQVRAITWK 420  
QY 419 KRPQNGIINQYRVKVLVPETGIIILENTLLTGNEVINDPMAPEIUNIVEPMVGLYEGSA 478  
DB 421 KRPQNGIINQYRVKVLVPETGIIILENTLLTGNE-INDPMAPEIUNIVEPMVGLYEGSA 479  
QY 479 EMSDDLHSLATFIYNHSHDPKFNARNAEDQTSVVTTRNOYITDIAAQLSVIRLVP 538  
DB 480 EMSDDLHSLATFIYNHSHDPKFNARNAEDQTSVVTTRNOYITDIAAQLSVIRLVP 539  
QY 539 FTEHMLSVSAPFTMGEGPPTVLSVTRQOVPSIKIINYKNISSSILLIYWDPEYPNGK 598  
DB 540 FTEHMLSVSAPFTMGEGPPTVLSVTRQOVPSIKIINYKNISSSILLIYWDPEYPNGK 599  
QY 599 ITHYTIYAMELDTNRAFOITIDNSFLIT--GLKYYTKYKMRVAASTHVGESSLSBENDI 656  
DB 600 ITHYTIYAMELDTNRAFOITIDNSFLITIGLKKYTKYKMRVAASTHVGESSLSBENDI 659  
QY 657 FVTSDEPSSQDVEIDVTADEIRLKWSPKENGIIIAEVLVYKNIIDITLYMKTST 716  
DB 660 FVTSDEPSSQDVEIDVTADEIRLKWSPKENGIIIAEVLVYKNIIDITLYMKTST 719  
QY 717 TDIIILANLRPHLTYNISVRSYTRFGHGNVSSLLSVRTSETVPSAPENITYKNISSGEI 776  
DB 720 TDIIILANLRPHLTYNISVRSYTRFGHGNVSSLLSVRTSETVPSAPENITYKNISSGEI 779  
QY 777 ELSFLPPSPNGIIKKYTYILKESNGNEERTINTTSLTONIKVLKKYQYIIIEVSASTLK 836  
DB 780 ELSFLPPSPNGIIKKYTYILKESNGNEERTINTTSLTONI--LKKYQYIIIEVSASTLK 837  
QY 837 GEGVRSAPISILTEEDAPSPPODFVKQSGVTVKLSWQPPLEPNGLIYYTYVYVW-NR 895  
DB 838 GEGVRSAPISILTEEDAPSPPODFVKQSGVTVKLSWQPPLEPNGLIYYTYVYVWRN 897  
QY 896 SSLKTNVTSLELSDLDVNVYSAAVTASTRFGGKTKGSNIISFQTPGAPSDPPKDV 955  
DB 898 SSLKTNVTSLELSDLDVNVYSAAVTASTRFGGKTKRSNIISFQTPG-PSDPPKDV 956  
QY 956 YYANLSSSSIIILFWTPPSKENGIIQYYSVYRNTSGTFMGNFTLHBLTNDFONMVTSTII 1015  
DB 957 YYANLSSSSIIILFWTPPSKENGIIQYYSVYRNTSGTFMGNFTLHBLTNDFONMVTSTII 1016  
QY 1016 DKLTIFSYTFMLTASTSVGNKSSDIIIEVYTDQDIPRGFVGNLYYESISSTAINVSW 1075  
DB 1017 DKLTIFSYTFMLTASTSVGNKSSDIIIEVYTDQDIPRGFVGNLYYESISSTAINVSW 1076  
QY 1076 PPAQNGLYVYVYSLIILOQTPRHVRPPLVYERSIYFDNLEKYTDVILKITSTEGKFS 1135  
DB 1077 PPAQNGLYVYVYSLIILOQTPRHVRPPLVYERSIYFDNLEKYTDVILKITSTEGKFS 1136

QY 1136 TYTAQLYIKTEEDVPETSPDIINTFKNLSSTSVLLSWDPPVKNGAIISVDLTLOGPNEY 1195  
DB 1137 TYTAQLYIKTEEDVPETSPDIINTFKNLSSTSVLLSWDPPVKNGAIISVDLTLOGPNEY 1196  
QY 1196 SPFTSDNYIIILBELSPFTLYSPFAAARTKGLGPSSILFPYTDDESVPAPPQNLTLINCT 1255  
DB 1197 SPFTSDNYIIILBELSPFTLYSPFAAARTKGLGPSSILFPYTDDESVPAPPQNLTLINCT 1256  
QY 1256 SPFWMLKWSPLPGGIVKYGKIHHEHETDTIYKNIISGFKTEAKLVGLEPVSYSIRV 1315  
DB 1257 SPFWMLKWSPLPGGIVKYGKIHHEHETDTIYKNIISGFKTEAKLVGLEPVSYSIRV 1316  
QY 1316 SAFTKVGNGNQSNVVKFTTQBSVDPVQNMOCMATSWOSVLVKWDPKKKANGIITQYV 1375  
DB 1317 SAFTKVGNGNQSNVVKFTTQBSVDPVQNMOCMATSWOSVLVKWDPKKKANGIITQYV 1376  
QY 1376 TVERNSTKVSPODHMYTFIKLLANTSYVFKVRASTASAGGDESTCHVSTLPETVPSVPTN 1435  
DB 1377 TVERNSTKVSPODHMYTFIKLLANTSYVFKVRASTASAGGDESTCHVSTLPETVPSVPTN 1436  
QY 1436 IAPSDVQSTSATLTWIRPDTIILGYFQNYKITTLQRAQCKEWESEECVYQKIQYLYEAH 1495  
DB 1437 IAPSDVQSTSATLTWIRPDTIILGYFQNYKITTLQRAQCKEWESEECVYQKIQYLYEAH 1496  
QY 1496 LTEETVYGLKRWTRFPQVAASTNAGYGNASNIWSTKTLPDPPDGPENVHVATSPFSI 1555  
DB 1497 LTEETVYGLKRWTRFPQVAASTNAGYGNASNIWSTKTLPDPPDGPENVHVATSPFSI 1556  
QY 1556 SLSWSEPAVITGPTCYLIDVKSVDNDEFNISFKSNEENKTIKOLEIFTRYSVVITAF 1615  
DB 1557 SLSWSEPAVITGPTCYLIDVKSVDNDEFNISFKSNEENKTIKOLEIFTRYSVVITAF 1616  
QY 1616 TGNISAAVYEGKSSAEMIVTTLESAPKOPNNMTFQKIPDEVTKFOLTLPSPQPNQNIQ 1675  
DB 1617 TGNISAAVYEGKSSAEMIVTTLESAPKOPNNMTFQKIPDEVTKFOLTLPSPQPNQNIQ 1676  
QY 1676 VYQALVYREDDPTAVQIHNLSIIQKNTNVIAMLEGLKGGHTYNIISVYAVNSAGAPKVP 1735  
DB 1677 VYQALVYREDDPTAVQIHNLSIIQKNTNVIAMLEGLKGGHTYNIISVYAVNSAGAPKVP 1736  
QY 1736 MRITWDIKAPAPKTKPTPIYDATGKLVSTTTITIRMPICVYDDHGPKNQVVLATET 1795  
DB 1737 MRITWDIKAPAPKTKPTPIYDATGKLVSTTTITIRMPICVYDDHGPKNQVVLATET 1796  
QY 1796 GAQHDGNTVTKYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYIIGADNACMIPGN 1855  
DB 1797 GAQHDGNTVTKYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYIIGADNACMIPGN 1856  
QY 1856 EDKICNGPLKPKKQVLFKPRATNIMGQFTDSYSDPVKTLGEGLSBRTVEIILSVTLCL 1915  
DB 1857 EDKICNGPLKPKKQVLFKPRATNIMGQFTDSYSDPVKTLGEGLSBRTVEIILSVTLCL 1916  
QY 1916 SIILLGTAIFAPARIRQKQEGGYSPQDAEIIDTKLKDQITVADLKBRLTR--- 1972  
DB 1917 SIILLGTAIFAPARIRQKQEGGYSPQDAEIIDTKLKDQITVADLKBRLTRLLS 1976  
QY 1973 -----PISKSGFLOHVEBELTNNNLKFOEESSELPKFLQDLSSSTDAIPLMNAKRRFPN 2026  
DB 1977 YRKSIPKISKSGFLOHVEBELTNNNLKFOEESSELPKFLQDLSSSTDAIPLMNAKRRFPN 2036  
QY 2027 IKPYNNNRVKLIADASVPGSYINASYISGYLCPNEFIATQGPLGTGDFWVWVETR 2086  
DB 2037 IKPY-NNNRVKLIADASVPGSYINASYISGYLCPNEFIATQGPLGTGDFWVWVETR 2095  
QY 2087 AKTLVMTQCPKGRIRCHOYWPEDKNKPVTVFGDVIITKLMEDVQIDWTIRDKIERHGD 2146  
DB 2096 AKTLVMTQCPKGRIRCHOYWPEDKNKPVTVFGDVIITKLMEDVQIDWTIRDKIERHGD 2155  
QY 2147 CMTVQCQNTAMPPEHGVSPENSAPLTHFVKLVASRAHDTTPMIHVSAGVGTGVFIAD 2206  
DB 2156 CMTVQCQNTAMPPEHGVSPENSAPLTHFVKLVASRAHDTTPMIHVSAGVGTGVFIAD 2215  
QY 2207 HLTQHINDHDFVDIYGLVAELSRBMRMCMVQNLAQYIFLHQCILDLNLSKSNQPCFVNY 2266



|          |                                       |   |      |  |    |  |
|----------|---------------------------------------|---|------|--|----|--|
| Db       | 2216                                  | HLTQHINDHFDYDIYGLVAELSERMCVQNLQAYIFLHQCILDLLSNKGSNPICFVNY | 2275 |  | FT | /note= "Potentially phosphorylated"              |
|          |                                       |   |      |  | FT | /note= "Potentially phosphorylated"              |
| Qy       | 2267                                  | SALQKWDSLDAMEGDSVELEWEETTM                                | 2291 |  | FT | /note= "Potentially phosphorylated"              |
|          |                                       |   |      |  | FT | /note= "Potentially phosphorylated"              |
| Db       | 2276                                  | SALQKWDSLDAMEGDSVELEWEETTM                                | 2300 |  | FT | /note= "Potentially phosphorylated"              |
|          |                                       |   |      |  | FT | /note= "Potentially phosphorylated"              |
| RESULT 6 |                                       |   |      |  | FT | /note= "Potentially phosphorylated"              |
| AAG79724 |                                       |   |      |  | FT | /label= Fibronectin type III domain              |
| ID       | AAG79724                              | standard; protein; 2299 AA.                               |      |  | FT | /note= "Identified using HMWER_PPFAM"            |
| XX       | AAG79724;                             |   |      |  | FT | /note= "Potentially glycosylated"                |
| XX       | DT                                    |   |      |  | FT | /note= "Potentially phosphorylated"              |
| XX       | XX                                    |   |      |  | FT | /note= "Potentially phosphorylated"              |
| XX       | 04-MAR-2003                           | (first entry)   |      |  | FT | /note= "Potentially phosphorylated"              |
| DE       | Human KPP-2,                          | Incyte ID No. 7480588CD1.                                 |      |  | FT | /note= "Potentially phosphorylated"              |
| XX       | XX                                    |   |      |  | FT | /note= "Potentially phosphorylated"              |
| KW       | Kinase; phosphatase; KPP;             | cell proliferation; arteriosclerosis;                     |      |  | FT | /note= "Potentially glycosylated"                |
| KW       | atherosclerosis; cirrhosis;           | hepatitis; reproduction; infertility;                     |      |  | FT | /note= "Potentially glycosylated"                |
| KW       | paroxysmal nocturnal haemoglobinuria; | polycythaemia vera; psoriasis;                            |      |  | FT | /note= "Potentially phosphorylated"              |
| KW       | primary thrombocytopaenia;            | cancer; development; renal tubular acidosis;              |      |  | FT | /note= "Potentially phosphorylated"              |
| KW       | anaemia; mental retardation;          | neurological disorder; Alzheimer's disease;               |      |  | FT | /note= "Potentially phosphorylated"              |
| KW       | Parkinson's disease; epilepsy;        | Sjogren's syndrome; uveitis; asthma;                      |      |  | FT | /note= "Potentially phosphorylated"              |
| KW       | menstrual cycle; autoimmune;          | inflammation; rheumatoid arthritis; AIDS;                 |      |  | FT | /note= "Potentially phosphorylated"              |
| KW       | autoimmune thyroiditis;               | contact dermatitis; Crohn's disease; allergy;             |      |  | FT | /note= "Potentially phosphorylated"              |
| KW       | diabetes mellitus;                    | glomerulonephritis; Goodpasture's syndrome; gout;         |      |  | FT | /note= "Potentially glycosylated"                |
| KW       | Graves' disease; Hashimoto's          | thyroiditis; irritable bowel syndrome;                    |      |  | FT | /note= "Potentially phosphorylated"              |
| KW       | multiple sclerosis;                   | osteoarthritis; osteoporosis; pancreatitis;               |      |  | FT | /note= "Potentially phosphorylated"              |
| KW       | Reiter's syndrome; rat;               | glomerular mesangial cell receptor protein;               |      |  | FT | /label= Fibronectin_type_III_domain              |
| KW       | tyrosine phosphatase precursor;       | enzyme.   |      |  | FT | /note= "Identified using HMWER_PPFAM"            |
| XX       | XX                                    |   |      |  | FT | /note= "Potentially phosphorylated"              |
| OS       | Homo sapiens.                         |   |      |  | FT | /note= "Potentially phosphorylated"              |
| XX       | XX                                    |   |      |  | FT | /note= "Potentially phosphorylated"              |
| FH       | Key                                   | Location/Qualifiers                                       |      |  | FT | /note= "Potentially phosphorylated"              |
| FT       | Peptide                               | 1..19   |      |  | FT | /label= Signal peptide                           |
| FT       | FT                                    | /label= Signal peptide                                    |      |  | FT | /note= "Identified using HMWER"                  |
| FT       | Peptide                               | 1..17   |      |  | FT | /label= Signal peptide                           |
| FT       | FT                                    | /label= Signal peptide                                    |      |  | FT | /note= "Identified using HMWER"                  |
| FT       | Modified-site                         | 17  |      |  | FT | /note= "Potentially phosphorylated"              |
| FT       | Modified-site                         | 28  |      |  | FT | /note= "Potentially phosphorylated"              |
| FT       | Domain                                | 51..138   |      |  | FT | /label= Fibronectin_type_III_domain              |
| FT       | FT                                    | /label= Fibronectin_type_III_domain                       |      |  | FT | /note= "Identified using HMWER_PPFAM"            |
| FT       | Protein                               | 65..91  |      |  | FT | /label= Receptor tyrosine Kinase class V_protein |
| FT       | FT                                    | /label= Receptor tyrosine Kinase class V_protein          |      |  | FT | /note= "Identified using BLIMPS_BLOCKS"          |
| FT       | Modified-site                         | 104   |      |  | FT | /note= "Potentially phosphorylated"              |
| FT       | Protein                               | 115..145  |      |  | FT | /label= Receptor tyrosine Kinase class V_protein |
| FT       | FT                                    | /label= Receptor tyrosine Kinase class V_protein          |      |  | FT | /note= "Identified using BLIMPS_BLOCKS"          |
| FT       | Modified-site                         | 120   |      |  | FT | /note= "Potentially phosphorylated"              |
| FT       | Domain                                | 145..205  |      |  | FT | /label= Glomerular mesangial_cell_receptor       |
| FT       | FT                                    | /label= Glomerular mesangial_cell_receptor                |      |  | FT | /note= "Identified using BLAST_PRODOR"           |
| FT       | Domain                                | 150..286  |      |  | FT | /label= Fibronectin_type_III_domain              |
| FT       | FT                                    | /label= Fibronectin_type_III_domain                       |      |  | FT | /note= "Identified using HMWER_PPFAM"            |
| FT       | Modified-site                         | 155   |      |  | FT | /note= "Potentially glycosylated"                |
| FT       | Modified-site                         | 162   |      |  | FT | /note= "Potentially glycosylated"                |
| FT       | Modified-site                         | 182   |      |  | FT | /note= "Potentially phosphorylated"              |
| FT       | Modified-site                         | 186   |      |  | FT | /note= "Potentially phosphorylated"              |
| FT       | Modified-site                         | 200   |      |  | FT | /note= "Potentially glycosylated"                |
| FT       | FT                                    |   |      |  | FT | /note= "Potentially glycosylated"                |

FT Modified-site 715 /note= "Potentially phosphorylated"  
 FT Modified-site 731 /note= "Potentially glycosylated"  
 FT Modified-site 733 /note= "Potentially phosphorylated"  
 FT Modified-site 751 /note= "Potentially phosphorylated"  
 FT Modified-site 757 /note= "Potentially phosphorylated"  
 FT Domain 759.842 /label= Fibronectin type III domain  
 FT /note= "Identified using HMMER\_Pfam"  
 FT Modified-site 761 /note= "Potentially phosphorylated"  
 FT Modified-site 765 /note= "Potentially glycosylated"  
 FT Modified-site 767 /note= "Potentially phosphorylated"  
 FT Modified-site 770 /note= "Potentially glycosylated"  
 FT Modified-site 772 /note= "Potentially phosphorylated"  
 FT Modified-site 809 /note= "Potentially glycosylated"  
 FT Modified-site 824 /note= "Potentially phosphorylated"  
 FT Modified-site 834 /note= "Potentially phosphorylated"  
 FT Modified-site 849 /note= "Potentially phosphorylated"  
 FT Domain 854.936 /label= Fibronectin type III domain  
 FT /note= "Identified using HMMER\_Pfam"  
 FT Modified-site 862 /note= "Potentially phosphorylated"  
 FT Modified-site 870 /note= "Potentially phosphorylated"  
 FT Modified-site 895 /note= "Potentially glycosylated"  
 FT Modified-site 898 /note= "Potentially phosphorylated"  
 FT Modified-site 903 /note= "Potentially glycosylated"  
 FT Modified-site 907 /note= "Potentially phosphorylated"  
 FT Modified-site 912 /note= "Potentially phosphorylated"  
 FT Modified-site 927 /note= "Potentially phosphorylated"  
 FT Domain 948.1040 /label= Fibronectin type III domain  
 FT /note= "Identified using HMMER\_Pfam"  
 FT Modified-site 959 /note= "Potentially glycosylated"  
 FT Modified-site 988 /note= "Potentially glycosylated"  
 FT Modified-site 996 /note= "Potentially glycosylated"  
 FT Modified-site 998 /note= "Potentially glycosylated"  
 FT Modified-site 1008 /note= "Potentially phosphorylated"  
 FT Modified-site 1013 /note= "Potentially glycosylated"

Query Match 98.4%; Score 11797; DB 6; Length 2299;

Best Local Similarity 98.5%; Pred. No. 0;

Matches 2272; Conservative 6; Mismatches 5; Indels 24; Gaps 8;

QY 1 MDPLIFLLFGTSETQVDNVNVPGETYDITISSISTYTSPTVRIVTPNVTKGPPV 60

DB 1 MDPLIFLLFGTSETQVDNVNVPGETYDITISSISTYTSPTVRI-----GBGPPV 55

QY 61 FLGERVGSAGILLSWNTPPNPNGBRIISYIVKYKVCVPMQVYTVQVRSPDSLEVLTTN 120  
 DB 56 FLGERVGSAGILLSWNTPPNPNGBRIISYIVKYKVCVPMQVYTVQVRSPDSLEVLTTN 115  
 QY 121 LNPGETTYEIKVAEENSAGIGVSDPFLFOTAESAPGKVVNLTVYANASAVKLIWYLPQ 180  
 DB 116 LNPGETTYEIKVAEENSAGIGVSDPFLFOTAESAPGKVVNLTVYANASAVKLIWYLPQ 175  
 QY 181 PNGKITSPKISVKHARSIGIVKDVSRVEDILTGLPECNENSESEFLMSTASPSPTLGRV 240  
 DB 176 PNGKITSPKISVKHARSIGIVKDVSRVEDILTGLPECNENSESEFLMSTASPSPTLGRV 235  
 QY 241 TTPSRTTHSSSTLTQNEISSVWKEPISFVWTHLRPTTYTLFVSAATTEAGYIDSTIVRT 300  
 DB 236 TTPSRTTHSSSTLTQNEISSVWKEPISFVWTHLRPTTYTLFVSAATTEAGYIDSTIVRT 295  
 QY 301 PESVPEGPQNCVTGNITGKSPSILWDPPPTIVTGFSRVVELYGPS--GRILDNSTKDLK 358  
 DB 296 PESVPEGPQNCVTGNITGKSPSILWDPPPTIVTGFSRVVELYGPSGAGRILDNSTKDLK 355  
 QY 359 PAFNTLTPTMYDVVYIAAETSAGTGPKGNISVFTPPDVGAVFDLQLAEVESTQVRAITWK 418  
 DB 356 PAFNTLTPTMYDVVYIAAETSAGTGPKGNISVFTPPDVGAVFDLQLAEVESTQVRAITWK 415  
 QY 419 KPRQNGIINQYRVKVLVPETGIIILENTLLTGNEYINDPMAPEIYVIVPEVGLYEGSA 478  
 DB 416 KPRQNGIINQYRVKVLVPETGIIILENTLLTGNE-INDPMAPEIYVIVPEVGLYEGSA 474  
 QY 479 EMSDDLHSLATPIYNSHPDKNPPARNRAEDQTSVVTTRNOYITDIAAQLSVIRRLVP 538  
 DB 475 EMSDDLHSLATPIYNSHPDKNPPARNRAEDQTSVVTTRNOYITDIAAQLSVIRRLVP 534  
 QY 539 FTEHMSVSAFTIMGEGPPTVLSVTRQOVPSIKIINYKNISSSSILLYWMPPEYPNGK 598  
 DB 535 FTEHMSVSAFTIMGEGPPTVLSVTRQOVPSIKIINYKNISSSSILLYWMPPEYPNGK 594  
 QY 599 ITHYTIYAMELDTNRAFOITITDINSFLITG----LKYTKYKVRVAASHTDSESSISEEN 654  
 DB 595 ITHYTIYAMELDTNRAFOITITDINSFLITGKQWLKKTYYKVRVAASHTDSESSISEEN 654  
 QY 655 DIFVTSSEDEPSSPDQVEVDIVTADERLKWSPPEKNGIILIAEVLKNIDTLYMKN 714  
 DB 655 DIFVTSSEDEPSSPDQVEVDIVTADERLKWSPPEKNGIILIAEVLKNIDTLYMKN 714  
 QY 715 STTDIILNLRPHLYNISVRSYTRFGHGNQVSSLSVRSVTSVTPDSPAENITYKNISSG 774  
 DB 715 STTDIILNLRPHLYNISVRSYTRFGHGNQVSSLSVRSVTSVTPDSPAENITYKNISSG 774  
 QY 775 EIELSFLPPSPNGIISKYTIYLRKSNGBERTINTSLTONIKVKLKKYQVYIEVSAST 834  
 DB 775 EIELSFLPPSPNGIISKYTIYLRKSNGBERTINTSLTONIKVKLKKYQVYIEVSAST 834  
 QY 835 LKGEVRSAPISILTEEDAPSPDQFSVKQLSGVTVKLSWQPPLEPNGIILYYTVYVW- 893  
 DB 835 LKGEVRSAPISILTEEDAPSPDQFSVKQLSGVTVKLSWQPPLEPNGIILYYTVYVW 894  
 QY 894 NRSSLKTNVTSLELSDLDYVVEYSAYVASTRFGDKTGSNIISFOTPEGASDPDK 953  
 DB 895 NRSSLKTNVTSLELSDLDYVVEYSAYVASTRFGDKTGSNIISFOTPEG-PSDPPK 953  
 QY 954 DVYANLSSSSIILFWTPPSKNGIIOYYSVYVYRNTSGTFMONTLHETNDFDNTVST 1013  
 DB 954 DVYANLSSSSIILFWTPPSKNGIIOYYSVYVYRNTSGTFMONTLHETNDFDNTVST 1013  
 QY 1014 IIDKLTIYSYTFWLTASTSVGNKSSDIIEVYTDODIPEGFVGNLYTESISSAINVS 1073  
 DB 1014 IIDKLTIYSYTFWLTASTSVGNKSSDIIEVYTDQDVPEGFVGNLYTESISSAINVS 1073  
 QY 1074 WPPAQNGLVYVYVSLILQOTPRHVRPPLVYERSIYFDNLEKTYDILKLTPTTEKGF 1133  
 DB 1074 WPPAQNGLVYVYVSLILQOTPRHVRPPLVYERSIYFDNLEKTYDILKLTPTTEKGF 1133  
 QY 1134 SDTYTAQLYIKTEEDVPETSPIINTFKNLSSVLLSWDPPVVKPNGAIISYDLTLQGPNE 1193

|||||  
1134 SDTYTAQLYKTEEDVPETSPINTPKNLSSVLLSWDPVPKNGAIIISYDLTLQGPNE 1193  
|||||  
1194 NYSFITSNDYIILEELSPPTLYSFFFAAARTRKGLGPSSILFPYTDSEVPLAPQNLTLIN 1253  
|||||  
1194 NYSFITSNDYIILEELSPPTLYSFFFAAARTRKGLGPSSILFPYTDSEVPLAPQNLTLIN 1253  
|||||  
1254 CTSDFVWLKWSPLPGGIVKVYSRKIHHEHETDIYKNIISGFKTEAKLVGLEPVSYSI 1313  
|||||  
1254 CTSDFVWLKWSPLPGGIVKVYSRKIHHEHETDIYKNIISGFKTEAKLVGLEPVSYSI 1313  
|||||  
1314 RVSFAFTKVGNGQFNSVNVKFTTQESVPDVQNMCMATSWQSVLVKWDPPKKANGIITQY 1373  
|||||  
1314 RVSFAFTKVGNGQFNSVNVKFTTQESVPDVQNMCMATSWQSVLVKWDPPKKANGIITQY 1373  
|||||  
1374 MVTVERNSTKVSQDHYTFIKLLANTSVYFKVRASTASAGEDESTCHVSTLPETVPSPV 1433  
|||||  
1374 MVTVERNSTKVSQDHYTFIKLLANTSVYFKVRASTASAGEDESTCHVSTLPETVPSPV 1433  
|||||  
1434 TNAFSDVOSTSATLTWIRPDITLGYFQNYKITTQRAOKCKEWESEECVEYKQIYLYE 1493  
|||||  
1434 TNAFSDVOSTSATLTWIRPDITLGYFQNYKITTQRAOKCKEWESEECVEYKQIYLYE 1493  
|||||  
1494 AHLTEETVYGLKKFRWYRQVAASSTNAGYGNASNMISTKTLPQPPDPPEVHVATSPF 1553  
|||||  
1494 AHLTEETVYGLKKFRWYRQVAASSTNAGYGNASNMISTKTLPQPPDPPEVHVATSPF 1553  
|||||  
1554 SISISWSEPAVITGPTCYLIDVKSVDNDFNIFSKSNEENKTIEIKOLEIFTRYSVVIT 1613  
|||||  
1554 SISISWSEPAVITGPTCYLIDVKSVDNDFNIFSKSNEENKTIEIKOLEIFTRYSVVIT 1613  
|||||  
1614 AFTGNISAAVVGKSAEMIITLESAPKDPNNMTFOKIPDEVTKFQTLFPPSPNGN 1673  
|||||  
1614 AFTGNISAAVVGKSAEMIITLESAPKDPNNMTFOKIPDEVTKFQTLFPPSPNGN 1673  
|||||  
1674 IQVYQALVYREDPTAVQIHNLSIIQKTNFTFIAMLEGLKGHTYNIISYAVNSAGAGPK 1733  
|||||  
1674 IQVYQALVYREDPTAVQIHNLSIIQKTNFTFIAMLEGLKGHTYNIISYAVNSAGAGPK 1733  
|||||  
1734 VPMRITMDIKAPARPKTPTIYDATGKLLVSTTITIRMPICYSDHGPINQVQLAT 1793  
|||||  
1734 VPMRITMDIKAPARPKTPTIYDATGKLLVSTTITIRMPICYSDHGPINQVQLAT 1793  
|||||  
1794 ETGAOHGDNVTWYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYIIGADNACMP 1853  
|||||  
1794 ETGAOHGDNVTWYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYIIGADNACMP 1853  
|||||  
1854 GNEDKICNGPLPKPKQYLPKFRATNIMGFTSDYSDPDKVTLGEGLSERTVEIILSVTL 1913  
|||||  
1854 GNEDKICNGPLPKPKQYLPKFRATNIMGFTSDYSDPDKVTLGEGLSERTVEIILSVTL 1913  
|||||  
1914 ILSIILLGTAIFAFARIQKQEGGYSPQDAEIIDTKLQDLITVADLELKDERLTR- 1972  
|||||  
1914 ILSIILLGTAIFAFARIQKQEGGYSPQDAEIIDTKLQDLITVADLELKDERLTR 1973  
|||||  
1973 -----PISKKSFLQVHEELCTNNNLKFOREFSELKPELODLSSTDADLPWNRKNFP 2024  
|||||  
1974 LSYRKSIFKPSKKSFLQVHEELCTNNNLKFOREFSELKPELODLSSTDADLPWNRKNFP 2033  
|||||  
2025 PNIKPYNNNRNVKLIADASVPSGDIYASVIGYLCNPNFIATQGPLGTVGDFWRWVWE 2084  
|||||  
2034 PNIKPY-NNNRVKNLIADASVPSGDIYASVIGYLCNPNFIATQGPLGTVGDFWRWVWE 2092  
|||||  
2085 TRAKTLVMLTQCPKGRIRCHQVWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERH 2144  
|||||  
2093 TRAKTLVMLTQCPKGRIRCHQVWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERH 2152  
|||||  
2145 GDCMTVRQCNFTAWPBGHPENSAPLIHFVKLVRSRAHDTTPMIVHCSAGVGRGTGVFTA 2204  
|||||  
2153 GDCMTVRQCNFTAWPBGHPENSAPLIHFVKLVRSRAHDTTPMIVHCSAGVGRGTGVFTA 2212  
|||||  
2205 LDHLTQHINDHFDVDIYGLVAELSRMCMQVNLQAQYIFLHQICILDLLSNKGSNQPICFV 2264  
|||||

Db 2213 LDHLTQHINDHFDVDIYGLVAELSRMCMQVNLQAQYIFLHQICILDLLSNKGSNQPICFV 2272  
QY 2265 NYSALQKQWSDLDAMEGDVELEWEETM 2291  
Db 2273 NYSALQKQWSDLDAMEGDVELEWEETM 2299  
RESULT 7  
AAO18736  
ID AAO18736 standard; protein; 2281 AA.  
XX  
AC AAO18736;  
XX  
DT 24-OCT-2002 (first entry)  
XX  
DE Human NOV2a protein.  
XX  
KW Human; NOVX; autoimmune disease; cancer; infection; inflammatory disease;  
KW storage disorder; muscle disorder; neurodegenerative disorder; hypotonic;  
KW developmental defect; neuroprotective; antiparkinsonian; hypotensive;  
KW hypertensive; haemostatic; cardiac; antiangiogenic; dermatological;  
KW immunosuppressive; antiinflammatory; virucide; antibacterial; anti-HIV;  
KW antiparasitic; antiallergic; antiasthmatic; antineoplastic; antidiabetic;  
KW nephrotropic; anorectic; antidiabetic; immunomodulator; antipsoriatic;  
KW antinfertility; antitumor; antidiabetic; cerebroprotective; anticonvulsant;  
KW tranquilizer; analgesic.  
XX  
OS Homo sapiens.  
OS WO200257450-A2.  
XX  
XX 25-JUL-2002.  
XX  
PF 29-NOV-2001; 2001WO-US048922.  
XX  
XX 29-NOV-2000; 2000US-0253834P.  
PR 30-NOV-2000; 2000US-0250926P.  
PR 25-JAN-2001; 2001US-0264180P.  
PR 20-AUG-2001; 2001US-0313658P.  
PR 05-OCT-2001; 2001US-0327456P.  
PR 28-NOV-2001; 2001US-00327456.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Edinger S, Macdougall JR, Millet I, Ellemann K, Stone DJ; Burgess CE;  
PI Gerlach V, Grosse WM, Alsobrook JP, Lepley DM, Rieger D, Mishra V;  
PI Casman SJ, Spytek KA, Boldog FL, Li L, Padigaru M, Mishra V;  
PI Patturajan M, Shenoy S, Rastelli L, Tchernev VT, Vernet CAM;  
PI Zerhusen BD, Malyankar UM, Guo X, Miller CE, Gangolli EA;  
XX  
XX WPI; 2002-590741/63.  
XX N-PSDB; APT06281.  
XX  
XX Novel isolated polypeptide, designated NOVX, useful for treating or  
XX preventing in NOVX-associated disorders e.g. cardiomyopathy,  
XX atherosclerosis, diabetes, cancer, allergy, asthma, Crohn's disease.  
XX  
XX Claim 1; Page 25-26; 353pp; English.  
XX  
XX The present invention provides the protein and coding sequences of  
XX several novel human proteins, designated NOVX. These can be used in the  
XX treatment of, amongst others, cancer, autoimmune diseases, infections,  
XX inflammatory diseases, storage disorders, muscle disorders, the present  
XX neurodegenerative diseases and developmental defects. The present  
XX sequence is a protein of the invention  
XX  
XX Sequence 2281 AA;  
SQ

Query Match 93.4%; Score 11191; DB 5; Length 2281;  
Best Local Similarity 94.5%; Pred. No. 0;  
Matches 2171; Conservative 40; Mismatches 63; Indels 24; Gaps 16;



Db 2144 FTAPWPHGVSPNSAPLIHFVFKLVRASRAHDTTPMIVHCAGVGRGTGVFALDHLTQHIND 2203  
Qy 2215 HDPVDIYGLVAELSRMCMQVNLAQYIFLHQICILDLLSNKGSNPICFVNTYSALQKMD 2274  
Db 2204 HDPVDIYGLVAELSRMCMQVNLAQYIFLHQICILDLLSNKGSNPICFVNTYSALQKMD 2263  
Qy 2275 LDAME-GDVELEWEETM 2291  
Db 2264 LDAMEGGDVELEWEETM 2281

RESULT 8  
ADH41619  
ID ADH41619 standard; protein; 2281 AA.  
XX  
AC ADH41619;  
XX  
XX  
25-MAR-2004 (first entry)  
XX  
XX Novel human protein NOV15d.  
XX  
XX cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;  
KW antidiabetic; immunosuppressive; anti-Hiv; neuroprotective; nootropic;  
KW antiparkinsonian; antiasthmatic; antiinfertility; cardiomyopathy;  
KW atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;  
KW multiple sclerosis; graft-versus-host disease; Alzheimer's disease;  
KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;  
KW tissue typing; preventive medicine; pharmacogenomic; vaccine.  
XX  
XX Homo sapiens.  
XX  
XX WO2003102159-A2.  
XX  
XX 11-DEC-2003.  
XX  
XX 04-JUN-2003; 2003WO-US017573.  
XX  
XX 04-JUN-2002; 2002US-0385490P.  
PR 04-JUN-2002; 2002US-0385615P.  
PR 04-JUN-2002; 2002US-0385755P.  
PR 05-JUN-2002; 2002US-0386041P.  
PR 06-JUN-2002; 2002US-0386355P.  
PR 06-JUN-2002; 2002US-0386357P.  
PR 06-JUN-2002; 2002US-0386447P.  
PR 06-JUN-2002; 2002US-0386459P.  
PR 06-JUN-2002; 2002US-0386465P.  
PR 06-JUN-2002; 2002US-0386864P.  
PR 07-JUN-2002; 2002US-0386701P.  
PR 07-JUN-2002; 2002US-0386796P.  
PR 07-JUN-2002; 2002US-0386931P.  
PR 07-JUN-2002; 2002US-0387078P.  
PR 07-JUN-2002; 2002US-0387081P.  
PR 07-JUN-2002; 2002US-0387083P.  
PR 10-JUN-2002; 2002US-0387429P.  
PR 10-JUN-2002; 2002US-0387540P.  
PR 10-JUN-2002; 2002US-0387866P.  
PR 10-JUN-2002; 2002US-0387866P.  
PR 11-JUN-2002; 2002US-0387606P.  
PR 11-JUN-2002; 2002US-0387610P.  
PR 11-JUN-2002; 2002US-0387659P.  
PR 11-JUN-2002; 2002US-0387668P.  
PR 11-JUN-2002; 2002US-0387696P.  
PR 11-JUN-2002; 2002US-0387859P.  
PR 12-JUN-2002; 2002US-0387934P.  
PR 12-JUN-2002; 2002US-0387960P.  
PR 12-JUN-2002; 2002US-0388022P.  
PR 12-JUN-2002; 2002US-0388096P.  
PR 12-JUN-2002; 2002US-0388432P.  
PR 12-JUN-2002; 2002US-0388479P.  
PR 12-JUN-2002; 2002US-0388479P.  
PR 13-JUN-2002; 2002US-0389123P.  
PR 14-JUN-2002; 2002US-0389120P.  
PR 14-JUN-2002; 2002US-0389146P.  
PR 17-JUN-2002; 2002US-0389742P.  
PR 18-JUN-2002; 2002US-0389604P.

PR 18-JUN-2002; 2002US-0389884P.  
PR 19-JUN-2002; 2002US-0390006P.  
PR 19-JUN-2002; 2002US-0390144P.  
PR 19-JUN-2002; 2002US-0390209P.  
PR 25-JUN-2002; 2002US-0391726P.  
PR 06-AUG-2002; 2002US-0401628P.  
PR 09-AUG-2002; 2002US-0402268P.  
PR 12-AUG-2002; 2002US-0402822P.  
PR 13-AUG-2002; 2002US-0403458P.  
PR 15-AUG-2002; 2002US-0403617P.  
PR 15-AUG-2002; 2002US-0403732P.  
PR 26-AUG-2002; 2002US-0406182P.  
PR 12-SEP-2002; 2002US-0410085P.  
PR 13-SEP-2002; 2002US-0410505P.  
PR 23-SEP-2002; 2002US-0412955P.  
PR 30-SEP-2002; 2002US-0415195P.  
PR 23-OCT-2002; 2002US-0420627P.  
PR 23-OCT-2002; 2002US-0420718P.  
PR 24-OCT-2002; 2002US-0420852P.  
PR 31-OCT-2002; 2002US-0422750P.  
PR 01-NOV-2002; 2002US-0423095P.  
PR 05-NOV-2002; 2002US-0423748P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;  
PI Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;  
PI Ellerman K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;  
PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;  
PI Khrantsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;  
PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CB;  
PI Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;  
PI Rieger DK, Rothenberg ME, Shenoy SG, Shinkets RA, Smithson G;  
PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;  
PI Wolenc AR, Zhong M, Zhong H;  
XX  
XX WPI: 2004-053467/05.  
DR N-PSDB; ADH41618.  
XX  
XX New NOVX polypeptides and nucleic acid molecules useful for preventing or  
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in  
PT pharmacogenomics.  
XX  
PS Claim 2; SEQ ID NO 172; 1503pp; English.  
XX  
XX The invention relates to 566 new isolated human polypeptides and their  
CC encoding genes, sequences that are at least 95% identical to these or  
CC sequences comprising one or more conservative substitutions in these. The  
CC polypeptide, polynucleotide and antibodies against the polypeptides are  
CC useful in diagnosing, treating or preventing NOVX-associated disorders,  
CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,  
CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,  
CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.  
CC The nucleic acids are further used as hybridization probes, in chromosome  
CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The  
CC polypeptides are also useful as vaccines. This sequence represents an  
CC example of the polypeptide of the invention.  
XX  
SQ Sequence 2281 AA;  
Query Match 93.4%; Score 11191; DB 8; Length 2281;  
Best Local Similarity 94.5%; Pred. No. 0;  
Matches 2171; Conservative 40; Mismatches 63; Indels 24; Gaps 16;  
Qy 1 MDPLIIFLLFIQTSETQVDVSNVVGTRDYDTISSISTYTPVTRIVPNVTKGPPV 60  
Db 1 MDPLIIFLLFIQTSETQVDVSNVVGTRDYDTISSISTYTPVTRIGASN--EGGPPV 58  
Qy 61 FLAGRVGSAGILLSNWTPPNPGRISIVIVKVCPCPMQVTVQVRKPSLEVLITN 120  
Db 59 FLAGRVGSAGILLSNWTPPNPGRISIVIVKVCPCPMQVTVQVRKPSLEVLITN 118

QY 121 LNPFTTVEIKVAEENSAGIVPSDPELFQTAESAPGKVNLTVYAYNASAVKLINWLPQ 180  
Db 119 LNPFTTVEIKVAEENSAGIVPSDPELFQTAESAPGKVNDFTGEAVPFS-KLMWYTSAT 177  
QY 181 PNGKITSFISVKHARSIGVWVDVSRVEDILTKLP-ECNENSESFLWSTASPSPTLGR 239  
Db 178 KK-KITSFKISVKHARSIGVWVEVSIRVECILSASPLHCNENSESFLWSTASPSPTLGR 236  
QY 240 VTPPSRTTHSSSTLTQNEISSVWKEPISFVVTTHLRYPTTYLFEVSAATTEAGYIDSTIYR 299  
Db 237 VTPPSRTTHSSSTLTQNEISSV-KEPISFVVTTHLRYPTTYLFEVSAATTEAGYIDSTIYR 295  
QY 300 TPESVPEGPNQCVTGNITGKSFILWDPPTIIVTGKFSYRVELYGPS-GRILDNSTKDLK 358  
Db 296 TPESVPEGPNQCVTGNITGKSFILWDPPTIIVTGKFSYRVELYGPSAGRILDNSTKDLK 355  
QY 359 FAFNTLTPFTWYDVYIAAETSAGTGPKSNISVFTPPDVGAVFDQLAEBVSTQVRITWK 418  
Db 356 FAFNTLTPFTWYDVYIAAETSAGTGPKSNISVFTPPDVGAVFDQLAEBVSTQVRITWK 415  
QY 419 KRPQNGIINQVRKVLVPETGIIILENTLLTGNEYINDMPAPEIYNIYVPMVGLYEGSA 478  
Db 416 KRPQNGIINQVRKVLVPETGIIILENTLLTGNE-INDMPAPEIYNIYVPMVGLYEGSA 474  
QY 479 EMSDDLHSLATFIYNSHPDKNPARNAEDOTSPVVTTRNOYITDIAAQLSVIRRLVP 538  
Db 475 EMSDDLHSLATFIYNSHPDKNPARNAEDOTSPVVTTRNOYITDIAAQLSVIRRLRR 534  
QY 539 FTEHMSVSAFTMGPPVNLVSRTRQOVPSIKIINX--NISSSILLWYDPEYEN 596  
Db 535 FWAETHGFRYTIMSS-----ASRNLTSPPGLSAQNFVRVTHVITEVFLHWDPDP--P 586  
QY 597 GKITHYIYAMELDT-NRAFOITIDN-SFLITGLKKYTKYKMRVAASVTHGESSISEEN 654  
Db 587 VFFHHVLIILDEVNQSKSILRLTNSLSVLIGLKKYTKYKMRVAASVTHGESSISEEN 646  
QY 655 DIFVRTSEDEPSSPODVEIDVTADEIRLKWSPPEKPNGLIIAYEVLKNDITLYMKN 714  
Db 647 DIFVRTSEDEPSSPODVEIDVTADEIRLKWSPPEKPNGLIIAYEVLKNDITLYMKN 706  
QY 715 STTDIILRLNRLHLYNISVRSYTRFEGHNOVSSLSVTSVTPDSPAENITYKNISSG 774  
Db 707 STTDIILRLNRLHLYNISVRSYTRFEGHNOVSSLSVTSVTPDSPAENITYKNISSG 766  
QY 775 EIELSLPSSPNGIILKTYILKRSNGNEERTINTSLTQNLKVLKKTQYIIIEVSAST 834  
Db 767 EIELSLPSSPNGIILKTYILKRSNGNEERTINTSLTQNLKVLKKTQYIIIEVSAST 826  
QY 835 LKGEVRSAPIISILTEEDAPSPQDFSVKQLSGVTVKLSWQPLEPNGIILYTYVYWN 894  
Db 827 LKGEVRSAPIISILTEEDAPSPQDFSVKQLSGVTVKLSWQPLEPNGIILYTYVYWN- 885  
QY 895 RSLKTIYNTETSELSLDLDYVVEYSAYTASTRFGDKTGKSNIIISFQTPPEGAPSPKX 954  
Db 886 RSLKTIYNTETSELSLDLDYVVEYSAYTASTRFGDKGKSNIIISFQTPPEG-PSDPPK 944  
QY 955 VYIANLSSSIIILFWTPPSKPNGLIOYYSVYTRNTSGTFMQNFTLHENDFDMTVSTI 1014  
Db 945 VYIANLSSSIIILFWTPPSKPNGLIOYYSVYTRNTSGTFMQNFTLHENDFDMTVSTI 1004  
QY 1015 IDKLATIFSYYTFLWTASTSVGNKSSDIIIEVYTDQDPEGFVGNLTYESISSTAINVSW 1074  
Db 1005 IDKLATIFSYYTFLWTASTSVGNKSSDIIIEVYTDQDPEGFVGNLTYESISSTAINVSW 1064  
QY 1075 VPPAQPNGLVFTVSVLILQOTPRHVRPPLVTVYERSIYFONLEKYTDYILKITPSTKGFPS 1134  
Db 1065 VPPAQPNGLVFTVSVLILQOTPRHVRPPLVTVYERSIYFONLEKYTDYILKITPSTKGFPS 1124  
QY 1135 DRYTAQLYIKTBEDVPETSPINTFNKLSSTVLSWDPKPNCAIISYDILTQGNEN 1194  
Db 1125 DRYTAQLYIKTBEDVPETSPINTFNKLSSTVLSWDPKPNCAIISYDILTQGNEN 1184  
QY 1195 YSFITSDNYIILBELSPFTLYSFFAAARTKGLGSSILFFYTDSESVLAPPQNLTLINC 1254

Db 1185 YSFITSDNYIILBELSPFTLYSFFAAARTKGLGSSILFFYTDSESVLAPPQNLTLINC 1244  
QY 1255 TSDFWLWKSPPGLPGIVKVSFKIHEHETDIYKNIISGPKTEAKLVGLPEVSYISIR 1314  
Db 1245 TSDFWLWKSPPGLPGIVKVSFKIHEHETDIYKNIISGPKTEAKLVGLPEVSYISIR 1304  
QY 1315 VSAFTKVGNGNQSNNVKTQESVDPVQVQNMCMATSWQSVLVKWDPPKANGIITQYM 1374  
Db 1305 VSAFTKVGNGNQSNNVKTQESVDPVQVQNMCMATSWQSVLVKWDPPKANGIITQYM 1364  
QY 1375 VTVERNSTKVSPODHYMTPIKLLANTSYVFKVRASTASAGDESTCHVSTLPEVSVPT 1434  
Db 1365 VTVERNSTKVSPODHYMTPIKLLANTSYVFKVRASTASAGDESTCHVSTLPEVSVPT 1424  
QY 1435 NIAPSDVOSTATLWIRPDTTILGYFQNYKIITQTLRAOCKEWESECEVEYQIKOYLVEA 1494  
Db 1425 NIAPSDVOSTATLWIRPDTTILGYFQNYKIITQTLRAOCKEWESECEVEYQIKOYLVEA 1484  
QY 1495 HUTBETVYGLKKFRWYRFQVAASTNAGYGNASNIWISTKTLPDPPGPPENHVAVTSPPS 1554  
Db 1485 HUTBETVYGLKKFRWYRFQVAASTNAGYGNASNIWISTKTLPDPPGPPENHVAVTSPPS 1544  
QY 1555 ISISWSEPAVITGPTCYLIDVKSVDNDEENISFIKSNEENKTIEIKDLIFTRYSVVIITA 1614  
Db 1545 ISISWSEPAVITGPTCYLIDVKSVDNDEENISFIKSNEENKTIEIKDLIFTRYSVVIITA 1604  
QY 1615 FTGNTSAAVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTQFOLTFLPSPQNGNI 1674  
Db 1605 FTGNTSAAVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTQFOLTFLPSPQNGNI 1664  
QY 1675 QYQALVYREDDPTAVQIHNLSIIQKTNTFVIALEGLKGHTYNIISVYAVNSAGAGPKV 1734  
Db 1665 QYQALVYREDDPTAVQIHNLSIIQKTNTFVIALEGLKGHTYNIISVYAVNSAGAGPKV 1724  
QY 1735 PMRITWIDKAPARPKTPTPIYDATGKLLVTSTTTITRMPICIYSDDHGPIKNVQVLATE 1794  
Db 1725 PMRITWIDKAPARPKTPTPIYDATGKLLVTSTTTITRMPICIYSDDHGPIKNVQVLATE 1784  
QY 1795 TGAQHDGNTWKYDAYFNKARPYFTNEGPNPCPTGKTKFSGNEEYIIIGADNACWIPG 1854  
Db 1785 TGAQHDGNTWKYDAYFNKARPYFTNEGPNPCPTGKTKFSGNEEYIIIGADNACWIPG 1844  
QY 1855 NEDKICNGPLPKKQYLFKFRATNIMGQFSDSDYSDPVKTLGEGLSERTVEIILSVTLCI 1914  
Db 1845 NEDKICNGPLPKKQYLFKFRATNIMGQFSDSDYSDPVKTLGEGLSERTVEIILSVTLCI 1904  
QY 1915 LSIILLGTAIFAFARIROKQEGGYTSPQDABIIDTKLKDQIITVADIJELKDERLTRPI 1974  
Db 1905 LSIILLGTAIFAFARIROKQEGGYTSPQDABIIDTKLKDQIITVADIJELKDERLTRPI 1964  
QY 1975 SKKSFLOHVEELCTNNLKFQEFSELKPELODLSSTADLPMNRAKORFPNIKPVNNNN 2034  
Db 1965 SKKSFLOHVEELCTNNLKFQEFSELKPELODLSSTADLPMNRAKORFPNIKPVNNNN 2023  
QY 2035 RVKLITADASVPGSDYINASYISGILCPNEFTATQCPGTPGVDFWRVWVETRAKTLVMLT 2094  
Db 2024 RVKLITADASVPGSDYINASYISGILCPNEFTATQCPGTPGVDFWRVWVETRAKTLVMLT 2083  
QY 2095 QCFEKGRIRCHQYWPEDNKPVTVFGDIIVITKLMEDVQIDWITRDLKIERHGDCMTVRQCN 2154  
Db 2084 QCFEKGRIRCHQYWPEDNKPVTVFGDIIVITKLMEDVQIDWITRDLKIERHGDCMTVRQCN 2143  
QY 2155 FTAWPEHGVSPENASPLIHFVKLVASRAHDTTPMIVHCSAGVGRGVGFIALDHLQIIND 2214  
Db 2144 FTAWPEHGVSPENASPLIHFVKLVASRAHDTTPMIVHCSAGVGRGVGFIALDHLQIIND 2203  
QY 2215 HDFVDIYGLVABLRSERCMQNLQAYIFLHQICILDLNKGSNOPICFVNYTSALQKXMS 2274  
Db 2204 HDFVDIYGLVABLRSERCMQNLQAYIFLHQICILDLNKGSNOPICFVNYTSALQKXMS 2263  
QY 2275 LDAME-GOVELEWSETTM 2291  
||||| |||||||

Db 2264 LDAMEGGDVELEWEETM 2281

RESULT 9

ID ADH41613

ADH41613 standard; protein; 2281 AA.

XX AC ADH41613;

XX DT 25-MAR-2004 (first entry)

XX DE Novel human protein NOV15a.

XX KW cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;

XX KW antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;

XX KW antiparkinsonian; antiasthmatic; antiinfertility; cardiomyopathy;

XX KW atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;

XX KW multiple sclerosis; graft-versus-host disease; Alzheimer's disease;

XX KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;

XX KW tissue typing; preventive medicine; pharmacogenomic; vaccine.

XX OS Homo sapiens.

XX PN WO2003102159-A2.

XX PD 11-DEC-2003.

XX PF 04-JUN-2003; 2003WO-US017573.

XX PR 04-JUN-2002; 2002US-0385490P.

XX PR 04-JUN-2002; 2002US-0385615P.

XX PR 04-JUN-2002; 2002US-0385755P.

XX PR 05-JUN-2002; 2002US-0386041P.

XX PR 06-JUN-2002; 2002US-0386355P.

XX PR 06-JUN-2002; 2002US-0386357P.

XX PR 06-JUN-2002; 2002US-0386447P.

XX PR 06-JUN-2002; 2002US-0386459P.

XX PR 06-JUN-2002; 2002US-0386465P.

XX PR 06-JUN-2002; 2002US-0386864P.

XX PR 07-JUN-2002; 2002US-0386701P.

XX PR 07-JUN-2002; 2002US-0386796P.

XX PR 07-JUN-2002; 2002US-0386931P.

XX PR 07-JUN-2002; 2002US-0387078P.

XX PR 07-JUN-2002; 2002US-0387081P.

XX PR 07-JUN-2002; 2002US-0387083P.

XX PR 10-JUN-2002; 2002US-0387429P.

XX PR 10-JUN-2002; 2002US-0387540P.

XX PR 10-JUN-2002; 2002US-0387866P.

XX PR 11-JUN-2002; 2002US-0387606P.

XX PR 11-JUN-2002; 2002US-0387610P.

XX PR 11-JUN-2002; 2002US-0387659P.

XX PR 11-JUN-2002; 2002US-0387668P.

XX PR 11-JUN-2002; 2002US-0387696P.

XX PR 11-JUN-2002; 2002US-0387859P.

XX PR 12-JUN-2002; 2002US-0387934P.

XX PR 12-JUN-2002; 2002US-0387960P.

XX PR 12-JUN-2002; 2002US-0388022P.

XX PR 12-JUN-2002; 2002US-0388096P.

XX PR 12-JUN-2002; 2002US-0388432P.

XX PR 12-JUN-2002; 2002US-0388479P.

XX PR 13-JUN-2002; 2002US-0389123P.

XX PR 14-JUN-2002; 2002US-0389120P.

XX PR 14-JUN-2002; 2002US-0389146P.

XX PR 17-JUN-2002; 2002US-0389742P.

XX PR 18-JUN-2002; 2002US-0389604P.

XX PR 18-JUN-2002; 2002US-0389884P.

XX PR 19-JUN-2002; 2002US-0390006P.

XX PR 19-JUN-2002; 2002US-0390144P.

XX PR 19-JUN-2002; 2002US-0390209P.

XX PR 25-JUN-2002; 2002US-0391726P.

XX PR 06-AUG-2002; 2002US-0401628P.

XX PR 09-AUG-2002; 2002US-0402268P.

XX PR 12-AUG-2002; 2002US-0402822P.

PR 13-AUG-2002; 2002US-0403458P.

PR 15-AUG-2002; 2002US-0403617P.

PR 15-AUG-2002; 2002US-0403732P.

PR 26-AUG-2002; 2002US-0406182P.

PR 12-SEP-2002; 2002US-0410085P.

PR 13-SEP-2002; 2002US-0410505P.

PR 20-SEP-2002; 2002US-0412955P.

PR 30-SEP-2002; 2002US-0415195P.

PR 23-OCT-2002; 2002US-0420627P.

PR 23-OCT-2002; 2002US-0420718P.

PR 24-OCT-2002; 2002US-0420852P.

PR 31-OCT-2002; 2002US-0422750P.

PR 01-NOV-2002; 2002US-0423095P.

PR 05-NOV-2002; 2002US-0423748P.

XX (CURA-) CURAGEN CORP.

XX Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;

PI Burgess CE, Casman SJ, Catterton E, Dhanabai M, Edinger SR;

PI Ellerman K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;

PI Grosse WM, Gunther E, Guo X, Gusev VY, Hermann JL, Ji W, Kekuda R;

PI Khramtsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;

PI MacLachlan T, Malyankar M, McQueeney K, Mezick AJ, Miller CE;

PI Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;

PI Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G;

PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;

PI Wolenc AR, Zhong M, Zhong H;

XX WPI; 2004-053467/05.

DR N-PSDB; ADH41612.

DR XX

XX New NOVX polypeptides and nucleic acid molecules useful for preventing or

PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,

PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in

PT pharmacogenomics.

XX Claim 2; SEQ ID NO 166; 1503pp; English.

XX The invention relates to 566 new isolated human polypeptides and their

CC encoding genes, sequences that are at least 95% identical to these or

CC sequences comprising one or more conservative substitutions in these. The

CC polypeptide, polynucleotide and antibodies against the polypeptides are

CC useful in diagnosing, treating or preventing NOVX-associated disorders,

CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,

CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,

CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.

CC The nucleic acids are further used as hybridization probes, in chromosome

CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The

CC polypeptides are also useful as vaccines. This sequence represents an

CC example of the polypeptide of the invention.

XX Sequence 2281 AA;

SQ

Query Match 93.4%; Score 11191; DB 8; Length 2281;

Best Local Similarity 94.5%; Pred. No. 0;

Matches 2171; Conservative 40; Mismatches 63; Indels 24; Gaps 16;

QY 1 MDFLIFLLFFIGTSETQVDVSNVPGTRYDITISSITTYTSPVTRIVTPNVTKEGPPV 60

Db 1 MDFLIFLLFFIGTSETQVDVSNVPGTRYDITISSITTYTSPVTRIVTPNVTKEGPPV 58

QY 61 FLAHERVGSAGILLSWNTPPNENGRIISYIVKYKEVCPMMQVTVTVQVRSKPDLSLEVLTN 120

Db 59 FLAHERVGSAGILLSWNTPPNENGRIISYIVKYKEVCPMMQVTVTVQVRSKPDLSLEVLTN 118

QY 121 LNPGGTYRIKVAENSAGIGVFSDFPLFQTAEASAFKGVNLTVEAYNASAVKLIWYLPQ 180

Db 119 LNPGGTYRIKVAENSAGIGVFSDFPLFQTAEASAFKGVNLTVEAYNASAVKLIWYLPQ 177

QY 181 PNGKITSPKISVKHARSIGIVKQVIRVEDILTGLP-ECNENSESEFLMSTASPSPTLGR 239

Db 178 KK-KITSPKISVKHARSIGIVKQVIRVEDILTGLP-ECNENSESEFLMSTASPSPTLGR 236



QY 240 VTPPSRTHSSSTLTQNEISSVWKEPISPVVTHLRPYTYTLFEVSAATTEAGYIDSTIVR 299  
DB 237 VTPPSRTHSSSTLTQNEISSV--KEPISFVTHLRPYTYTLFEVSAATTEAGYIDSTIVR 295  
QY 300 TPESVPEGPPONCVGNIGTKSPSILWDPTIIVTGKFSVRVELYGPS--GRILDNSTKDLK 358  
DB 296 TPESVPEGPPONCVGNIGTKSPSILWDPTIIVTGKFSVRVELYGPSAGRILDNSTKDLK 355  
QY 359 FAFNTNTPPTMYDVYIAAETSAGTGPKNISVFTPPDPGAVPDLQLAEVESQVRAITWK 418  
DB 356 FAFNTNTPPTMYDVYIAAETSAGTGPKNISVFTPPDPGAVPDLQLAEVESQVRAITWK 415  
QY 419 KRPQNGIINQYRVKVLVPETGIIILENTLLTGNEVINDPMAPEIIVNIPWVGLYESSA 478  
DB 416 KRPQNGIINQYRVKVLVPETGIIILENTLLTGNE--INDPMAPEIIVNIPWVGLYESSA 474  
QY 479 EMSDDLHSLATFIYNHSHPKNPPARNRAEDQTSFVVTTRNOYITDIAAQOLSXVIRRLVP 538  
DB 475 EMSDDLHSLATFIYNHSHPKNPPARNRAEDQTSFVVTTRNOYITDIAAQOLTYVLRRLR 534  
QY 539 FTEHMIISVAFPTIMGEGPPPTVLSVTRQOVPPSIIKINYK--NISSSILLWDPPEYPN 596  
DB 535 FWAETWGFGRYTIMSS-----ASRDNLTSPGLSAQNFRVTHVITTEVFLHWDPPD--P 586  
QY 597 GKITHYTIYAMELDT--NRAFOITIDN--SPLITGLKKYTKYKRVAASTHDGESSISEEN 654  
DB 587 VPFHHYLIITILDVENQSKSIILATLMSLSLVGLKKYTKYKRVAASTHVGESSISEEN 646  
QY 655 DIFVRISEBEPSSPODVEIDVTADEIRLKWSPPEKPNGLIIAYEVLKYNIDTLWKMT 714  
DB 647 DIFVRISEBEPSSPODVEIDVTADEIRLKWSPPEKPNGLIIAYEVLKYNIDTLWKMT 706  
QY 715 STTDIILRLNRLPHTLYNISVRSYTRFGHGNQVSSLSVRTSEVPDSAPENITYKNISSG 774  
DB 707 STTDIILRLNRLPHTLYNISVRSYTRFGHGNQVSSLSVRTSEVPDSAPENITYKNISSG 766  
QY 775 EIBLSFLPSSPNGIIKKYTIYLRSGNHEERTINTSLTONIKVLKKTQYHIEVSAST 834  
DB 767 EIBLSFLPSSPNGIIKKYTIYLRSGNHEERTINTSLTONIKGLKKTQYHIEVSAST 826  
QY 835 LKGEGRVSAPISILTEEDAPDSPQDPFSVKQLSGVTVKLSWQPLPENGIIILYTYVWN 894  
DB 827 LKGEGRVSAPISILTEEDAPDSPQDPFSVKQLSGVTVKLSWQPLPENGIIILYTYVWN-- 885  
QY 895 RSLKKTINTVETLSLESLDLDYNYEYSAVYTASTRFGDGKTSNIIISFQTPEGAPSDPPKD 954  
DB 886 RSLKKTINTVETLSLESLDLDYNYEYSAVYTASTRFGDGKTSNIIISFQTPEG--PSDPPKD 944  
QY 955 VYVANISSSIIILFWTPPKPNGIIQYISVYVYRNTSGTFMONTLHENDFDMNTVSTI 1014  
DB 945 VYVANISSSIIILFWTPPKPNGIIQYISVYVYRNTSGTFMONTLHEVTNDFDMNTVSTI 1004  
QY 1015 IDKLTIFSYTYFWLTASTSVGNKSSDIIIEVYTDODIPEGFVGNLTYESISTAINVSM 1074  
DB 1005 IDKLTIFSYTYFWLTASTSVGNKSSDIIIEVYTDODIPEGFVGNLTYESISTAINVSM 1064  
QY 1075 VPPAQPNGLVYVYVLSIILOQTPHVRPPLVYTERSIYFONLEKYTDYILKITPSTBKGPS 1134  
DB 1065 VPPAQPNGLVYVYVLSIILOQTPHVRPPLVYTERSIYFONLEKYTDYILKITPSTBKGPS 1124  
QY 1135 DTYTAQLYKTEEDVPETSPFIINTFKNLSTSVLLSWDPPVPKNGAIISYDLTLQGNEN 1194  
DB 1125 DTYTAQLYKTEEDVPETSPFIINTFKNLSTSVLLSWDPPVPKNGAIISYDLTLQGNEN 1184  
QY 1195 YSPITSDNYIILELSPFTLYSFAAARTKGLGPSIILFFYTDERSVPLAPQNLTLINC 1254  
DB 1185 YSPITSDNYIILELSPFTLYSFAAARTKGLGPSIILFFYTDERSVPLAPQNLTLINC 1244  
QY 1255 TSDPVLKWSPSPLPGGIVKVSFKIHEHTDIYIYKNISGFKTEAKLVGLEBVSYSIR 1314  
DB 1245 TSDPVLKWSPSPLPGGIVKVSFKIHEHTDIYIYKNISGFKTEAKLVGLEBVSYSIR 1304  
QY 1315 VSAFTKVGNGNQFSNVVKTTOESVDPVQVNMCMQMATSWQSVLVKWDPPKANGIITQYM 1374

RESULT 10  
ABP60058  
ID ABP60058 standard; protein; 2301 AA.  
XX  
AC ABP60058;

DB 1305 VSAFTKVGNGNQFSNVVKTTOESVDPVQVNMCMQMATSWQSVLVKWDPPKANGIITQYM 1364  
QY 1375 VTVERNSTKVSQDQHMHTFIKLLANTSYVFKVRASTASAGEDESTCHVSTLTPETVPSVPT 1434  
DB 1365 VTVERNSTKVSQDQHMHTFIKLLANTSYVFKVRASTASAGEDESTCHVSTLTPETVPSVPT 1424  
QY 1435 NIAPSDVQSTATLTIWRPDTTILGFQNYKIITQLRAQCKKEWSEBECVEYQIKIQLYLEA 1494  
DB 1425 NIAPSDVQSTATLTIWRPDTTILGFQNYKIITQLRAQCKKEWSEBECVEYQIKIQLYLEA 1484  
QY 1495 HUTEETVYGLKKFRVYRFOVAASTNAGYGNASNMWISTKTLPDPPDPPNHHVAVATSPS 1554  
DB 1485 HUTEETVYGLKKFRVYRFOVAASTNAGYGNASNMWISTKTLPDPPDPPNHHVAVATSPS 1544  
QY 1555 ISISWSEPAVITGPTCYLIDVKSVDNDEFNIFIKSNEENKTIIEIKDLBIFRYSVVIITA 1614  
DB 1545 ISISWSEPAVITGPTCYLIDVKSVDNDEFNIFIKSNEENKTIIEIKDLBIFRYSVVIITA 1604  
QY 1615 FTGNISAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKFQLTFLPPSQPNGNI 1674  
DB 1605 FTGNISAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKFQLTFLPPSQPNGNI 1664  
QY 1675 QYQALVYREDDPTAVQIHNLISIIOKTNTFVIAMLEGLKGGHTYNIISVAVANSAGAGPKV 1734  
DB 1665 QYQALVYREDDPTAVQIHNLISIIOKTNTFVIAMLEGLKGGHTYNIISVAVANSAGAGPKV 1724  
QY 1735 PMRIITMDIKAPARPKTKPTPIYDATGKLVATSTTTITIRMPICYISDDHGPINKNVVLATE 1794  
DB 1725 PMRIITMDIKAPARPKTKPTPIYDATGKLVATSTTTITIRMPICYISDDHGPINKNVVLATE 1784  
QY 1795 TGAQHDGNVTKWYDAYFNKARPYFTNEGPPNPCTEGTKFPGSNEBIYIIGADNACMIIPG 1854  
DB 1785 TGAQHDGNVTKWYDAYFNKARPYFTNEGPPNPCTEGTKFPGSNEBIYIIGADNACMIIPG 1844  
QY 1855 NEDKICNGPLKPKQYLFKFRATNIMGQFTSDYSDPVKTLGEGLSERTVEIILSVTLCI 1914  
DB 1845 NEDKICNGPLKPKQYLFKFRATNIMGQFTSDYSDPVKTLGEGLSERTVEIILSVTLCI 1904  
QY 1915 LSIIILLGTAFAPARIQKQEGGTVSPQDAIIDIITKLKLDQIITVADILEKDERLTRPI 1974  
DB 1905 LSIIILLGTAFAPARIQKQEGGTVSPQDAIIDIITKLKLDQIITVADILEKDERLTRPI 1964  
QY 1975 SKKSFLOHVEELCTNNLKFQBEFSELPKFLODLSSTDADLPNNRAKNRFPNIIKPYNNNN 2034  
DB 1965 SKKSFLOHVEELCTNNLKFQBEFSELPKFLODLSSTDADLPNNRAKNRFPNIIKPY--NNN 2023  
QY 2035 RVKLIADASVPGSDVINASYISGYLCPNEFIATQGPLGTVGDFWMMWETRAKTLVMLT 2094  
DB 2024 RVKLIADASVPGSDVINASYISGYLCPNEFIATQGPLGTVGDFWMMWETRAKTLVMLT 2083  
QY 2095 QCFEKGRIRCHQYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDQMTVRQCN 2154  
DB 2084 QCFEKGRIRCHQYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDQMTVRQCN 2143  
QY 2155 FTAPWEGHGVENSAPLIHFVKLVIRASRAHDTTTPMIVHCSAGVGTGTFVIALDHLTOHIND 2214  
DB 2144 FTAPWEGHGVENSAPLIHFVKLVIRASRAHDTTTPMIVHCSAGVGTGTFVIALDHLTOHIND 2203  
QY 2215 HDFVDIYGLVABLRSERMCMQNLQAYIFLHCILDLNLSKNSQNPICFVNTYSALQKQMS 2274  
DB 2204 HDFVDIYGLVABLRSERMCMQNLQAYIFLHCILDLNLSKNSQNPICFVNTYSALQKQMS 2263  
QY 2275 LDAME--GDVELEWEETTM 2291  
DB 2264 LDAMEGGDVELEWEETTM 2281

|    |   |     |      |   |      |
|----|---|-----|------|---|------|
| XX | 21-MAR-2003 (first entry)   | Db  | 241  | VTPTVTRTTQSSSTAARSKISSVWKBPISFVVVTHLRYTTLFEVSAVTEAGYIDSTIVR   | 300  |
| XX | Human phosphatase related protein #SEQ ID 4.                              | Qy  | 300  | TPESVPEGPONCVTGNITGKSFSLWDPPTIVTGKFSVRVELYGPSGRILDNSTKDLKF    | 359  |
| XX | Human phosphatase; enzyme; chromosome 12; colon adenocarcinoma;           | Db  | 301  | TPESVPEGPONCINGNVTKAFSLWDPPTIVTGKFSVRVELYGPSGRILDNSTKDLRF     | 360  |
| XX | placenta; gene therapy.   | Qy  | 360  | AFTNLTPFTMYDVYIAAETSAGTGPKSNISVFTPPDVPFCAVFDLQLAEBEVSTQVRIWKK | 419  |
| XX | Homo sapiens.   | Db  | 361  | AFTHLTPTFTMYDVYVAAETSAGVFKSNLSVFTPPDVPFCAVFDLQLAEBEATEIRITWRK | 420  |
| XX | WO200279452-A2.   | Qy  | 420  | PRQNGIINQVRKVLVPETGIILNTLLTGNEVINDPMAPEIVNVEPMVGLTEGSAB       | 479  |
| XX | 10-OCT-2002.  | Db  | 421  | PRQNGIISQVRKVSULETGVLVLENTLLTGDEISINPMSPIMNLVDPMTGFEYSGE      | 480  |
| XX | 02-APR-2002; 2002WO-US009992.   | Qy  | 480  | MSSDLHSLATFLYNSHPDKNFARNRADOETSPVVTTRNOYITDITAAQOLSVVIRRLVPP  | 539  |
| XX | 02-APR-2001; 2001US-00822871.   | Db  | 481  | MSSDLHSPASFLYNSHPHNDPPASTRAEEQSPVVTTRNQWMTDITABQOLSVVIRRLVPP  | 540  |
| XX | (PEKE ) PE CORP NY.   | Qy  | 540  | TEHMISVSAPFTMGEGPPTVLVSVTRQOVPSIKIINYKNISSSSSILLYWDPPEYNGKI   | 599  |
| XX | Webster M, Wei M, Di Francesco V, Beasley EM;                             | Db  | 541  | TEHTISVSAPFTMGEGPPTVLVSVTRQOVPSIIQIINYKNISSSSSILLYWDPPEYNGKI  | 600  |
| XX | WPI; 2003-046809/04.  | Qy  | 600  | THYTIYAMELDTNRAFQITTDINSFLITGLKKYTKYKRVAASTHGDGSSSISEENDIFVR  | 659  |
| XX | New human phosphatase peptides and nucleic acids encoding the peptides,   | Db  | 601  | THYTIYATELDTNRAFQITTDINSFLITGLKKYTKYKRVAASTHGVGSSSISEENDIFVR  | 660  |
| XX | useful as models in developing human therapeutic targets, in identifying  | Qy  | 660  | TSEDEPSSPDQVEIDVTADIEIRLKWSPPEKPGIITAYEVLVKNIDTLYMKNSTTDDI    | 719  |
| XX | therapeutic proteins or modulators of protein activity, and in gene       | Db  | 661  | TPEDEPSSPDQVQVTGVSPELRKWSPPPEKPGIITAYEVLVQNAUTLFLVKNSTTDDI    | 720  |
| XX | therapy.  | Qy  | 720  | ILRNLRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSGEIELS   | 779  |
| XX | Disclosure; Page 216-220; 220pp; English.                                 | Db  | 721  | IISDLKPYTLYNISIRSYTRLGHGNQSSLLSVRTSETVPSAPENITYKNISSGEIELS    | 780  |
| XX | The invention relates to an isolated human phosphatase peptide. The human | Qy  | 780  | FLPPSPNGIITKYTIYIKRNGNEERINTTSLTQNTKVLKKYTOYIIEVSASTLKGBG     | 839  |
| XX | phosphatase peptides and nucleic acids are useful as models for the       | Db  | 781  | FLPPSPNGIITKYTIYIKRNSNSHEARTINTTSLTQTIGGLKKYTHYVIEVSASTLKGBG  | 840  |
| XX | development of human therapeutic targets and in the identification of     | Qy  | 840  | VRSAPISTLTREDAPDSPQPSVKQLSGVTVKLSMQPPLPENGIIILYVTVVNRSLUK     | 899  |
| XX | therapeutic proteins. They also serve as targets for the development of   | Db  | 841  | INSRPSILTREDAPDSPQNPFSVKQLSGVTVMLSQPPLEPENGIIILYVTVVMDKSSUR   | 900  |
| XX | human therapeutic agents that modulate phosphatase activity in cells and  | Qy  | 900  | TINVTETSLLESLDLNVEYSAVYASTFRFGDKTGSNIISFQTPPEGAPSDPPKDVYAN    | 959  |
| XX | tissues that express the phosphatase. Experimental data indicates that    | Db  | 901  | AINATEASLVLSDLNDVYDYGACVTASTFRDGNARSSIIINFRTEGEPSPDPNDVHYVN   | 960  |
| XX | the cDNA is expressed in colon adenocarcinoma and placenta. The proteins  | Qy  | 960  | LSSSIIILFWTPSPKNGIIQYISVYVNTSGTFMONTLHETLNDFNMTVSTTIDKLT      | 1019 |
| XX | can be used to raise or to elicit another immune response, as a reagent   | Db  | 961  | LSSSIIILFWTPPPVXKNGIIQYISVYVYQNTSGTFVQNFLLQVTKESDNVTSARIYRLA  | 1020 |
| XX | in assays designed to determine the levels of the protein in biological   | Qy  | 1020 | IFSYTTFWLTASTSVGNKSSDIIEVVTDDIPEGFVGNLTYSISSTAINVSVPPAQ       | 1079 |
| XX | fluids, as markers for tissue in which the corresponding protein is       | Db  | 1021 | IFSYTTFWLTASTSVGNKSSDIIEVVTDDIPEGFVGNLTYSISSTALHVSWEPPSQ      | 1080 |
| XX | preferentially expressed, in the identification of modulators of the      | Qy  | 1080 | ENGLVYFVYSILQOT-PRHVRPPLVTVYERSIYFDNLEKYTDYIILKTPSTEKGFSDTYT  | 1138 |
| XX | peptides, and in pharmacogenomic analysis. The polypeptides and           | Db  | 1081 | ENGLVYFVYLSNLQOSPRHMTPLVTVYENSIDFDLEKYTDYIFKIPSTEKGFSEYTV     | 1140 |
| XX | polynucleotides may be used in gene therapy, and as antisense             | Qy  | 1139 | AQLYKTBEDVPETSPINTFNKLSSTSVLSWDPVPKNGAIISYDLTLQPNENYSPFI      | 1198 |
| XX | constructs to control phosphatase gene expression in cells, tissue and    | Db  | 1141 | TQLHKTBEDVPDTPPIINTFNKLSSTSVLSWDPPLKNGAILGYHLTLQGFHANHTFV     | 1200 |
| XX | organisms. The gene encoding the phosphatase of the invention is located  | Qy  | 1199 | TSDNVILILEELSPFTLYSPFAAARTRKGLGPSSILFEYTDSEVPPLAPQNLTLINCTSD  | 1258 |
| XX | on human chromosome 12. The current sequence represents a human           | Db  | 1201 | TSGNHVILEELSPFTLYSFFFAAARTMKGLGPSSILFFYTDSEAPLAPQNLTLINVTSD   | 1260 |
| XX | phosphatase protein related to the one of the invention                   | Qy  | 1259 | VMLKWSPLPGGIIVKYVSFKIHEHETDIYVYKNISGFKTEAKVLGLEPVSITYSRVSAP   | 1318 |
| XX | Sequence 2301 AA;   | Db  | 1261 | VMLTWSPLPGGIIVKYVSFKIHEHETDVFYKNISGLQTDAKLGLGLEPVSITYSVSAP    | 1320 |
| XX | Query Match 87.9%; Score 10536; DB 6; Length 2301;                        | Qy  | 1319 | TKVGNQFQSVNVKFTTQESVDPVQNMCCNATSWQSVLVKWDPPPKANGIITQWMTVE     | 1378 |
| XX | Best Local Similarity 85.9%; Pred. No. 0;                                 | Db  | 1321 | TKVGNQFQSVNVVEFTTQESVPEAVRNIECVARDWQSVSVRWDPPPKNGIITQWMTVE    | 1380 |
| XX | Matches 1977; Conservative 162; Mismatches 151; Indels 12; Gaps 4;        | Qy  |      |   |      |
| Qy | 1 MDPLIFLLFLTGTSQDVDSNVVPGTRYDITISSIS--TTYTSPVTRIVTNPVKPGPP               | 59  |      |   |      |
| Db | 1 MDHFPSFLLLTGTSQDVDSGDFGTGDTLSSVSATTYSSPVRFLATNVTXKPGPP                  | 60  |      |   |      |
| Qy | 60 VFLAGRVGSAGILLSSWNTPPNPNGRISYIVKYKEVCPWMQTVYTVQVRKPDLSLEVLIT           | 119 |      |   |      |
| Db | 61 VFLAGRVGSAGILLSSWNTPPNPNGRISYVVKYKEVCPWMQTVYTRARAKPDSLEVLIT            | 120 |      |   |      |
| Qy | 120 NLNPGTTYIKVAENASAGIVFSDPFLQTAESAPGKVNLITVEAYNASAVKLWYLP               | 179 |      |   |      |
| Db | 121 NLNPGTTYIKVAENASAGIVFSDPFLQTAESAPGKVNLITVEALNYSVNLWYLP                | 180 |      |   |      |
| Qy | 180 QNGKITSFKISVKHARSIGVVKDVSIRVEDILTGLPECNNSBSFLWSTASPSPTLGR             | 239 |      |   |      |
| Db | 181 QNGKITSFKISVKHARSIGVVKDVSIRVEDILSGKLPECNNSBSFLWSTASPSPTLGR            | 240 |      |   |      |
| Qy | 240 VTTPPSRTHSSSTLTQNEISSVWKEPISFVVTTHLRYTTLFEVSAATTEAGYIDSTIVR           | 299 |      |   |      |

```
QY 1379 RNSTKVSPOHMTYFKILLANTSYVFKVASTASAGBDEBCHVSTLPTVPSVPTNIAF 1438
DB 1381 GNSTKVSPPDPTTFTKLLPNTSYVFEVRASTASAGBESRCDISTLPTVPSAPTNAF 1440
QY 1439 SDVQSTSATLWIRPDITLGYFQNYKITTLQRAQCKWESECEVEYQKIQVLYEAHLTE 1498
DB 1441 SNVQSTSATLWTKPDITIGYFQNYKITTLQRAQCKWESECEVEYQKIQVLYEAHLTE 1500
QY 1499 ETYVGLKPRWRFRQVAASTNAGYGNASNWI STKTLPGPPDPPENHVHVAATSPFSISIS 1558
DB 1501 ETYVGLKPRWRFRQVAASTNAGYGNASNWI STKTLPGPPDPPENHVHVAATSPFSISIS 1560
QY 1559 WSEPAVITGPTCVLIDVKSVDNDEFNIFSIKNEENKTTIEIKDLLEIFTYSVVIATFQGN 1618
DB 1561 WSEPAVITGPTCVLIDVKSVDNDEFNIFSIKNEENKTTIEIKDLLEIFTYSVVIATFQGN 1620
QY 1619 ISAAVYBGGKSAEMIVTTLESAPKPPNNMTQKIPDEVTKEQLTFLPSPQNGNIQVQ 1678
DB 1621 VSRAYTDGKSSAEVIIITLESVPKPPNNMTQKIPDEVTKEQLTFLPSPQNGNIQVQ 1680
QY 1679 ALVYREDDPTAVQIHNLSIIQKNTNFVIAMLBGLKGHTYNI SVYAVNSAGAGPKVPMRI 1738
DB 1681 ALVYREDDPTAVQIHNLSIIQKNTNFVIAMLBGLKGHTYNI SVYAVNSAGAGPKVPMRI 1740
QY 1739 TMDIKAPARKPTPIYDATGKLLVTSITIRMPICYSDHGHGPKRVQVLATETGAQ 1798
DB 1741 TMDIKAPARKPTPIYDATGKLLVTSITIRMPICYSDHGHGPKRVQVLATETGAQ 1800
QY 1799 HGNVTKWDAVFNKARPYFTNEGPNPCTGKTFKFSNEEYIIGADNACMIQNEBK 1858
DB 1801 QDGNVTKWDAVFNKARPYFTNEGPNPCTGKTFKFSNEEYIIGADNACMIQNEBK 1860
QY 1859 ICNGPLKPKQVLFKFRATNIMGQTDSDYSPDKTLGGLSERTVEIILSVTLCLISII 1918
DB 1861 ICNGPLKPKQVLFKFRATNIMGQTDSDYSPDKTLGGLSERTVEIILSVTLCLISII 1920
QY 1919 LIGTAIFAPARIRQKQEGTYSPODAEIIDTKLKLQDITVADLELKDRLTR----- 1972
DB 1921 LIGTAIFAPARIRQKQEGTYSPODAEIIDTKLKLQDITVADLELKDRLTR----- 1980
QY 1973 ----PISKSPLOHVEELCTNNIKQEESSELPKFLQDLSSTDAIPLWNRKRNPNIKP 2029
DB 1981 STKPSKSPLOHVEELCTNNIKQEESSELPKFLQDLSSTDAIPLWNRKRNPNIKP 2040
QY 2030 YNNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQCPGTVGDFRVMWVETRAKT 2089
DB 2041 YNNNRVKLIADVSLPGSDYINASYISGYLCPNEFIATQCPGTVGDFRVMWVETRAKT 2099
QY 2090 LVMLTQCPEKGRIRCHQYWPEDNKPVTVFQDIVITKLMEDVQIDWITRDLKIERHGDGMT 2149
DB 2100 LVMLTQCPEKGRIRCHQYWPEDNKPVTVFQDIVITKLMEDVQIDWITRDLKIERHGDGMT 2159
QY 2150 VRQCNFTAMPHEGVSPNSAPLHFVKLVRASRAHDTTMI VCSAGVGRGTGFIALDHLT 2209
DB 2160 VRQCNFTAMPHEGVSPNTTFLHFVKLVRTSRAHDTTTPMVVHCSAGVGRGTGFIALDHLT 2219
QY 2210 QHINHDHFDYIYGLVAELSERCMVQNLQAOYIFLHQCTILDLSNKGNOPTCFVNYNSAL 2269
DB 2220 QHINHDHFDYIYGLVAELSERCMVQNLQAOYIFLHQCTILDLSNKGNOPTCFVNYNSAL 2279
QY 2270 QKMDSLDAMEGDVLEWEETTM 2291
DB 2280 QKMDSLDAMEGDVLEWEETTM 2301
```

RESULT 11

AAE37322

ID AAE37322 standard; protein; 1959 AA.

XX AAE37322;

AC AAE37322;

XX DT 07-AUG-2003 (first entry)

XX DE Human receptor tyrosine kinase protein #3.  
XX Human; receptor tyrosine phosphatase; diabetes; obesity; CNS disorder;  
KW multiple sclerosis; epilepsy; chronic obstructive pulmonary disease;  
KW COPD; Parkinson's disease; stroke; cardiovascular disorder; arrhythmia;  
KW congestive heart failure; myocardial infarction; ischaemic heart disease;  
KW gene therapy; anorectic; cardiant; neuroprotective; anticonvulsant;  
KW cerebroprotective; vasotropic; antiarrhythmic; receptor; enzyme.  
XX Homo sapiens.  
OS WO2003033688-A1.  
PN 24-APR-2003.  
PD 14-OCT-2002; 2002WO-EP011473.  
PF 16-OCT-2001; 2001US-0329329P.  
PR (FARB ) BAYER AG.  
PA Liou J;  
XX WPI: 2003-403215/38.  
XX N-PSDB; AAD56413.  
XX Novel polynucleotides encoding human receptor tyrosine phosphatase  
PT polypeptides, useful for treating diabetes, CNS disorders, obesity,  
PT chronic obstructive pulmonary disease and cardiovascular disorders.  
XX Disclosure; Fig 5; 163pp; English.  
XX The present invention relates to receptor tyrosine phosphatase proteins  
CC and polynucleotides encoding them. Sequences of the invention are useful  
CC in the preparation of medicaments for modulating the activity of receptor  
CC tyrosine phosphatase in disease such as diabetes, obesity, CNS disorders  
CC (multiple sclerosis, epilepsy, Parkinson's disease and stroke), chronic  
CC obstructive pulmonary disease (COPD) and cardiovascular disorders (e.g.  
CC congestive heart failure, myocardial infarction, ischaemic heart disease  
CC and arrhythmia). They are also used in gene therapy. The present sequence  
CC is human receptor tyrosine kinase protein  
XX Sequence 1959 AA;  
SQ

Query Match 82.4%; Score 9874; DB 6; Length 1959;

Best Local Similarity 96.0%; Pred. No. 0;

Matches 1998; Conservative 15; Mismatches 29; Indels 36; Gaps 4;

QY 331 IVTKFYSYRVELYGPSRI-----LDNSTKDLKFAFTNLTPFTWVDVYIAETSAGT 382

DB 1 LLGGKLTNRKDIHTKNPSVHHHQRPKVDKTK-----MGKKQSRKT 42

QY 333 GPKSNISVFTPPDVCAPVDLQLAESTVQVRIITWKKRQPNQIINQVRKVLVETGII 442

DB 43 GNSKQSTSPPPKQGRAVFDLQLAESTVQVRIITWKKRQPNQIINQVRKVLVETGII 102

QY 443 LENTLLTGNNEINDMPAPEIIVIEPMVGLYEGSAEMSSDLHSLATFIYNHDPKQVPA 502

DB 103 LENTLLTGNNEINDMPAPEIIVIEPMVGLYEGSAEMSSDLHSLATFIYNHDPKQVPA 162

QY 503 RNRAEDQTSPPVVTTRNQYITDIAEQLSVYIRRLVPFTTEHMSVSFAFTMGSPPTVLSV 562

DB 163 RNRAEDQTSPPVVTTRNQYITDIAEQLSVYIRRLVPFTTEHMSVSFAFTMGSPPTVLSV 222

QY 563 RTRQOVPSIKIINTKNISSSILLYWDPPEPNQKIHTYTIYAMELDTNRAFIQTIDN 622

DB 223 RTRQOVPSIKIINTKNISSSILLYWDPPEPNQKIHTYTIYAMELDTNRAFIQTIDN 282

QY 623 SFLITGLKKYTKYKMRVAASTHDGESSLSEENDIFVRTSEDEPSSPODVEVIDVTADSI 682

DB 283 SFLITGLKKYTKYKMRVAASTHDGESSLSEENDIFVRTSEDEPSSPODVEVIDVTADSI 342



PR 07-JUN-2002; 2002US-0386796P.  
 PR 07-JUN-2002; 2002US-0386931P.  
 PR 07-JUN-2002; 2002US-0387078P.  
 PR 07-JUN-2002; 2002US-0387081P.  
 PR 07-JUN-2002; 2002US-0387083P.  
 PR 10-JUN-2002; 2002US-0387429P.  
 PR 10-JUN-2002; 2002US-0387440P.  
 PR 10-JUN-2002; 2002US-0387866P.  
 PR 11-JUN-2002; 2002US-0387606P.  
 PR 11-JUN-2002; 2002US-0387610P.  
 PR 11-JUN-2002; 2002US-0387659P.  
 PR 11-JUN-2002; 2002US-0387688P.  
 PR 11-JUN-2002; 2002US-0387896P.  
 PR 11-JUN-2002; 2002US-0387859P.  
 PR 12-JUN-2002; 2002US-0387934P.  
 PR 12-JUN-2002; 2002US-0387960P.  
 PR 12-JUN-2002; 2002US-0388022P.  
 PR 12-JUN-2002; 2002US-0388096P.  
 PR 12-JUN-2002; 2002US-0388432P.  
 PR 13-JUN-2002; 2002US-0389123P.  
 PR 14-JUN-2002; 2002US-0389146P.  
 PR 17-JUN-2002; 2002US-0389742P.  
 PR 18-JUN-2002; 2002US-0389604P.  
 PR 18-JUN-2002; 2002US-0389884P.  
 PR 19-JUN-2002; 2002US-0390006P.  
 PR 19-JUN-2002; 2002US-0390144P.  
 PR 19-JUN-2002; 2002US-0390209P.  
 PR 25-JUN-2002; 2002US-0391726P.  
 PR 06-AUG-2002; 2002US-0401628P.  
 PR 09-AUG-2002; 2002US-0402268P.  
 PR 12-AUG-2002; 2002US-0402822P.  
 PR 13-AUG-2002; 2002US-0403458P.  
 PR 15-AUG-2002; 2002US-0403617P.  
 PR 26-AUG-2002; 2002US-0403732P.  
 PR 26-AUG-2002; 2002US-0406182P.  
 PR 12-SEP-2002; 2002US-0410085P.  
 PR 13-SEP-2002; 2002US-0410505P.  
 PR 23-SEP-2002; 2002US-0412955P.  
 PR 30-SEP-2002; 2002US-0415195P.  
 PR 23-OCT-2002; 2002US-0420627P.  
 PR 23-OCT-2002; 2002US-0420718P.  
 PR 24-OCT-2002; 2002US-0420852P.  
 PR 31-OCT-2002; 2002US-0422750P.  
 PR 01-NOV-2002; 2002US-0423095P.  
 PR 05-NOV-2002; 2002US-0423748P.

(CURA-) CURAGEN CORP.

PI Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;  
 PI Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;  
 PI Ellerman K, Ettenberg S, Gangoli EA, Gerlach VL, Gorman LJ;  
 PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;  
 PI Khrantsov NV, Larocheville WJ, Li L, Liang H, Low K, Macdougall JR;  
 PI MacLachlan T, Malvankar UM, McQueeney K, Mezick AJ, Miller CE;  
 PI Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;  
 PI Rieger DK, Rothenberg ME, Shenoy SG, Shinkets RA, Smithson G;  
 PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;  
 PI Wolenc AR, Zhong M, Zhong H;

WPI: 2004-053467/05.  
 N-PSDB; ADH41620.

PT New NOVX polypeptides and nucleic acid molecules useful for preventing or  
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
 PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in  
 PT pharmacogenomics.

XX Claim 2; SEQ ID NO 174; 1503pp; English.

PS The invention relates to 566 new isolated human polypeptides and their  
 CC encoding genes, sequences that are at least 95% identical to these or

CC sequences comprising one or more conservative substitutions in these. The  
 CC polypeptide, polynucleotide and antibodies against the polypeptides are  
 CC useful in diagnosing, treating or preventing NOVX-associated disorders,  
 CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,  
 CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,  
 CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.  
 CC The nucleic acids are further used as hybridization probes, in chromosome  
 CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The  
 CC polypeptides are also useful as vaccines. This sequence represents an  
 CC example of the polypeptide of the invention.

XX Sequence 909 AA;

Query Match 38.9%; Score 4661; DB 8; Length 909;  
 Best Local Similarity 99.6%; Pred. No. 1.9e-272;  
 Matches 905; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 18 QVDVSNVPGTRDYDTISSITVTSPVTRIVTNTVTKGPPVFLAGERVGSAGILLSWN 77  
 DB 1 QVDVSNVPGTRDYDTISSITVTSPVTRIVTNTVTKGPPVFLAGERVGSAGILLSWN 60  
 QY 78 TTPNPNGRISIVIKYKVCPPMOTVTVQVRKSPDSLEVLTLNLPNGTYYEIKVAAENSA 137  
 DB 61 TTPNPNGRISIVIKYKVCPPMOTVTVQVRKSPDSLEVLTLNLPNGTYYEIKVAAENSA 120  
 QY 138 GIGVFSDDPFLQTAESAPGKVNLTVEAYNASAVKLIWYLPQPNKGKITSFKISVKHARS 197  
 DB 121 GIGVFSDDPFLQTAESAPGKVNLTVEAYNASAVKLIWYLPQPNKGKITSFKISVKHARS 180  
 QY 198 GIWVKDVSIRVEDILTGLKPECNENSESEFLWSTASPSPTLGRVTPPSRTHSSSTLTQNE 257  
 DB 181 GIWVKDVSIRVEDILTGLKPECNENSESEFLWSTASPSPTLGRVTPPSRTHSSSTLTQNE 240  
 QY 258 ISSVMKEPISFVTHLRPVTTYLFEVSAATTEAGYIDSTIVRTPESVPSPGPPQNCVTGNI 317  
 DB 241 ISSVMKEPISFVTHLRPVTTYLFEVSAATTEAGYIDSTIVRTPESVPSPGPPQNCVTGNI 300  
 QY 318 TGKSFSLWDPTTIVTGKFSYRVELYGPSGRILDNSTKDLKFAFTNLTPFTYDVIYAAE 377  
 DB 301 TGKSFSLWDPTTIVTGKFSYRVELYGPSGRILDNSTKDLKFAFTNLTPFTYDVIYAAE 360  
 QY 378 TSAGTGPKSNISVFTPPDPVPGAVFDLQLAEVSTQVRIITWKKPQPNQINGIINQVRKVLVP 437  
 DB 361 TSAGTGPKSNISVFTPPDPVPGAVFDLQLAEVSTQVRIITWKKPQPNQINGIINQVRKVLVP 420  
 QY 438 ETGIIILENTLLTGNNIYINDPNAPEIVNIVEPMVGLYEGSAEMSSDLHSLATFIYNSHPD 497  
 DB 421 ETGIIILENTLLTGNNIYINDPNAPEIVNIVEPMVGLYEGSAEMSSDLHSLATFIYNSHPD 480  
 QY 498 KNFPAARNRAEDQTSFVVTTRNQYITDIAAEQLSYVIRRLVPPFTEHNMISVSAPFTIMGEGPP 557  
 DB 481 KNFPAARNRAEDQTSFVVTTRNQYITDIAAEQLSYVIRRLVPPFTEHNMISVSAPFTIMGEGPP 540  
 QY 558 TVLSVTRTQOVPSKIKINYNKNISSSILLYWDPEYPNGKITHYTIYAMELDTNRAFI 617  
 DB 541 TVLSVTRTQOVPSKIKINYNKNISSSILLYWDPEYPNGKITHYTIYAMELDTNRAFI 600  
 QY 618 TTIDNSFLITGLKKYTKYKMRVAASTHDCGESSISENDIFVRTSEDEPSSPODVVIDV 677  
 DB 601 TTIDNSFLITGLKKYTKYKMRVAASTHDCGESSISENDIFVRTSEDEPSSPODVVIDV 660  
 QY 678 TADEIRLKWSPPEKPNGIIIAYEVLYKNIDTLYMKNSTTIDILRLNRLPHTLYNISVRSY 737  
 DB 661 TADEIRLKWSPPEKPNGIIIAYEVLYKNIDTLYMKNSTTIDILRLNRLPHTLYNISVRSY 720  
 QY 738 TRFHGHNQVSSLLSVRTSETVPSAPENITYKNISGEIELSFLPPSSPNGIINKYTIYL 797  
 DB 721 TRFHGHNQVSSLLSVRTSETVPSAPENITYKNISGEIELSFLPPSSPNGIINKYTIYL 780  
 QY 798 KRSNGNEERTINTTSLTQNIKVKYQYIIIEVSASTLKGEGVRSAPISILTEEDAPDSP 857  
 DB 781 KRSNGNEERTINTTSLTQNIKVKYQYIIIEVSASTLKGEGVRSAPISILTEEDAPDSP 840



KW cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;  
KW antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;  
KW antiparkinsonian; antischismatic; antiinfertility; cardiomyopathy;  
KW atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;  
KW multiple sclerosis; graft-versus-host disease; Alzheimer's disease;  
KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;  
KW tissue typing; preventive medicine; pharmacogenomic; vaccine.

XX Homo sapiens.

XX WO2003102159-A2.

XX 11-DEC-2003.

XX 04-JUN-2003; 2003WO-US017573.

XX 04-JUN-2002; 2002US-0385490P.

XX 04-JUN-2002; 2002US-0385615P.

XX 04-JUN-2002; 2002US-0385755P.

XX 05-JUN-2002; 2002US-0386041P.

XX 06-JUN-2002; 2002US-0386355P.

XX 06-JUN-2002; 2002US-0386357P.

XX 06-JUN-2002; 2002US-0386447P.

XX 06-JUN-2002; 2002US-0386459P.

XX 06-JUN-2002; 2002US-0386465P.

XX 06-JUN-2002; 2002US-0386864P.

XX 07-JUN-2002; 2002US-0386701P.

XX 07-JUN-2002; 2002US-0386796P.

XX 07-JUN-2002; 2002US-0386931P.

XX 07-JUN-2002; 2002US-0387078P.

XX 07-JUN-2002; 2002US-0387081P.

XX 07-JUN-2002; 2002US-0387083P.

XX 10-JUN-2002; 2002US-0387429P.

XX 10-JUN-2002; 2002US-0387540P.

XX 10-JUN-2002; 2002US-0387866P.

XX 11-JUN-2002; 2002US-0387806P.

XX 11-JUN-2002; 2002US-0387810P.

XX 11-JUN-2002; 2002US-0387659P.

XX 11-JUN-2002; 2002US-0387668P.

XX 11-JUN-2002; 2002US-0387696P.

XX 11-JUN-2002; 2002US-0387859P.

XX 12-JUN-2002; 2002US-0387934P.

XX 12-JUN-2002; 2002US-0387960P.

XX 12-JUN-2002; 2002US-0388022P.

XX 12-JUN-2002; 2002US-0388096P.

XX 12-JUN-2002; 2002US-0388432P.

XX 12-JUN-2002; 2002US-0388479P.

XX 13-JUN-2002; 2002US-0389123P.

XX 14-JUN-2002; 2002US-0389120P.

PR

XX

PA

XX

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

05-NOV-2002; 2002US-0423748P.

(CURA-) CURAGEN CORP.

Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;

Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;

Ellerman K, Stettenberg S, Gangolli EA, Gerlach VL, Gorman L;

Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;

Khrantsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;

MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;

Milliet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Raetelli L;

Rieger DK, Rothenberg MB, Shenoy SG, Shinkets RA, Smithson G;

Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;

Wolenc AR, Zhong M, Zhong H;

WPI; 2004-053467/05.

N-PSDB; ADH41622.

New NOVX polypeptides and nucleic acid molecules useful for preventing or

treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,

atherosclerosis or diabetes, in chromosome mapping, tissue typing or in

pharmacogenomics.

Claim 2; SEQ ID NO 176; 1503pp; English.

The invention relates to 566 new isolated human polypeptides and their

encoding genes, sequences that are at least 95% identical to these or

sequences comprising one or more conservative substitutions in these. The

polypeptide, polynucleotide and antibodies against the polypeptides are

useful in diagnosing, treating or preventing NOVX-associated disorders,

e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,

diabetes, AIDS, multiple sclerosis, graft-versus-host disease,

Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.

The nucleic acids are further used as hybridization probes, in chromosome

mapping, tissue typing, preventive medicine, and pharmacogenomics. The

polypeptides are also useful as vaccines. This sequence represents an

example of the polypeptide of the invention.

Sequence 851 AA;

Query Match 37.4%; Score 4482; DB 8; Length 851;

Best Local Similarity 99.8%; Pred. No. 1.1e-261; Mismatches 2; Indels 0; Gaps 0;

Matches 849; Conservative 0;

1053 PEGFVGNLTYESISSTAINVSWPPAQNGLVYVYVLSILQQTPRHVRPLVYERSIYF 1112

1 PEGFVGNLTYESISSTAINVSWPPAQNGLVYVYVLSILQQTPRHVRPLVYERSIYF 60

1113 DNLEKYTDYILKTPSTKGFSDTYTAQLYIKTEEDVPSTPIINTFKNLSSTVLLSWD 1172

61 DNLEKYTDYILKTPSTKGFSDTYTAQLYIKTEEDVPSTPIINTFKNLSSTVLLSWD 120

1173 PPKNGGAIISYDLTLQGNENYSFTSDNYIILBELSPFTLYSFAAARTRKGLGPSSI 1232

121 PPKNGGAIISYDLTLQGNENYSFTSDNYIILBELSPFTLYSFAAARTRKGLGPSSI 180

1233 LFFYTDSEVPLAPPQNLTINCTDFVWLKNSPPLGPGIVKVSFKIHEHEDTIIYKN 1292

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db



|           |             |   |      |
|-----------|-------------|---|------|
| Db        | 421         | KCKEWESECVYQIKYLYEAHLTEETVYGLKFRWRFOVAASNAGYGNASWISTK                     | 480  |
| Qy        | 1533        | TLPGGPDGPPENHVAVTSFSSISWSBPVITGPTCYLIDVKSVDNDEFNISFIKSNE                  | 1592 |
| Db        | 481         | TLPGGPDGPPENHVAVTSFSSISWSBPVITGPTCYLIDVKSVDNDEFNISFIKSNE                  | 540  |
| Qy        | 1593        | ENKTIKIDLEIFTRYSVVITAFITGNISAAVYVEGSSAEMIVTTLESAPKDPNNMTFOK               | 1652 |
| Db        | 541         | ENKTIKIDLEIFTRYSVVITAFITGNISAAVYVEGSSAEMIVTTLESAPKDPNNMTFOK               | 600  |
| Qy        | 1653        | IDPEVTKFOLTPLPSQPNNGNIQVQALVYREDDPTAVQIHNLISIIQKTNFTVIAMLEGL              | 1712 |
| Db        | 601         | IDPEVTKFOLTPLPSQPNNGNIQVQALVYREDDPTAVQIHNLISIIQKTNFTVIAMLEGL              | 660  |
| Qy        | 1713        | KGGHTYNTISVYAVNSAGAGPKVPMRITWMDIKAPAPKTKPTPIYDATGKLLVTSITIR               | 1772 |
| Db        | 661         | KGGHTYNTISVYAVNSAGAGPKVPMRITWMDIKAPAPKTKPTPIYDATGKLLVTSITIR               | 720  |
| Qy        | 1773        | MEICYSDHGPINKVQVLATETGAQHDGNVTKWDYAFNKARPYFTNKGFPNPPCTEGK                 | 1832 |
| Db        | 721         | MEICYSDHGPINKVQVLATETGAQHDGNVTKWDYAFNKARPYFTNKGFPNPPCTEGK                 | 780  |
| Qy        | 1833        | TKPSGNEEYIIIGADNACMIPGNEKICNGPLKPKKQYLFKPRATNIMGQFTDSYSDPV                | 1892 |
| Db        | 781         | TKPSGNEEYIIIGADNACMIPGNEKICNGPLKPKKQYLFKPRATNIMGQFTDSYSDPV                | 840  |
| Qy        | 1893        | KTLGEGLSERT   | 1903 |
| Db        | 841         | KTLGEGLSERT   | 851  |
| RESULT 15 |             |   |      |
| ID        | ADH41617    | standard; protein; 740 AA.  |      |
| XX        | AC          | ADH41617;   |      |
| XX        | AC          | ADH41617;   |      |
| DT        | 25-MAR-2004 | (first entry)   |      |
| XX        | DE          | Novel human protein NOV15c.   |      |
| KW        |             | cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic; |      |
| KW        |             | antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;    |      |
| KW        |             | anti-parkinsonian; antiaesthetic; antifertility; cardiomyopathy;          |      |
| KW        |             | atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;           |      |
| KW        |             | multiple sclerosis; graft-versus-host disease; Alzheimer's disease;       |      |
| KW        |             | Parkinson's disease; asthma; fertility disorder; chromosome mapping;      |      |
| KW        |             | tissue typing; preventive medicine; pharmacogenomic; vaccine.             |      |
| OS        |             | Homo sapiens.   |      |
| XX        | XX          | WO2003102159-A2.  |      |
| XX        | XX          | 11-DEC-2003.  |      |
| XX        | XX          | 04-JUN-2003; 2003WO-US017573.   |      |
| XX        | XX          | 04-JUN-2003; 2002US-0385490P.   |      |
| XX        | XX          | 04-JUN-2003; 2002US-0385615P.   |      |
| XX        | XX          | 04-JUN-2003; 2002US-0385755P.   |      |
| XX        | XX          | 05-JUN-2003; 2002US-0386041P.   |      |
| XX        | XX          | 06-JUN-2003; 2002US-0386355P.   |      |
| XX        | XX          | 06-JUN-2003; 2002US-0386357P.   |      |
| XX        | XX          | 06-JUN-2003; 2002US-0386447P.   |      |
| XX        | XX          | 06-JUN-2003; 2002US-0386459P.   |      |
| XX        | XX          | 06-JUN-2003; 2002US-0386465P.   |      |
| XX        | XX          | 07-JUN-2003; 2002US-0386701P.   |      |
| XX        | XX          | 07-JUN-2003; 2002US-0386796P.   |      |
| XX        | XX          | 07-JUN-2003; 2002US-0386931P.   |      |
| XX        | XX          | 07-JUN-2003; 2002US-0387078P.   |      |
| XX        | XX          | 07-JUN-2003; 2002US-0387081P.   |      |

|    |  |                  |
|----|--|------------------|
| PR | 07-JUN-2002;   | 2002US-0387083P. |
| PR | 10-JUN-2002;   | 2002US-0387429P. |
| PR | 10-JUN-2002;   | 2002US-0387540P. |
| PR | 10-JUN-2002;   | 2002US-0387866P. |
| PR | 11-JUN-2002;   | 2002US-0387866P. |
| PR | 11-JUN-2002;   | 2002US-0387606P. |
| PR | 11-JUN-2002;   | 2002US-0387610P. |
| PR | 11-JUN-2002;   | 2002US-0387659P. |
| PR | 11-JUN-2002;   | 2002US-0387668P. |
| PR | 11-JUN-2002;   | 2002US-0387696P. |
| PR | 11-JUN-2002;   | 2002US-0387859P. |
| PR | 12-JUN-2002;   | 2002US-0387934P. |
| PR | 12-JUN-2002;   | 2002US-0387960P. |
| PR | 12-JUN-2002;   | 2002US-0388022P. |
| PR | 12-JUN-2002;   | 2002US-0388096P. |
| PR | 12-JUN-2002;   | 2002US-0388432P. |
| PR | 12-JUN-2002;   | 2002US-0388479P. |
| PR | 13-JUN-2002;   | 2002US-0389123P. |
| PR | 14-JUN-2002;   | 2002US-0389120P. |
| PR | 14-JUN-2002;   | 2002US-0389146P. |
| PR | 17-JUN-2002;   | 2002US-0389742P. |
| PR | 18-JUN-2002;   | 2002US-0389604P. |
| PR | 18-JUN-2002;   | 2002US-0389884P. |
| PR | 19-JUN-2002;   | 2002US-0390006P. |
| PR | 19-JUN-2002;   | 2002US-0390144P. |
| PR | 19-JUN-2002;   | 2002US-0390209P. |
| PR | 25-JUN-2002;   | 2002US-0391726P. |
| PR | 06-AUG-2002;   | 2002US-0401628P. |
| PR | 09-AUG-2002;   | 2002US-0402268P. |
| PR | 12-AUG-2002;   | 2002US-0402822P. |
| PR | 13-AUG-2002;   | 2002US-0403458P. |
| PR | 15-AUG-2002;   | 2002US-0403617P. |
| PR | 15-AUG-2002;   | 2002US-0403732P. |
| PR | 26-AUG-2002;   | 2002US-0406182P. |
| PR | 12-SEP-2002;   | 2002US-0410085P. |
| PR | 13-SEP-2002;   | 2002US-0410505P. |
| PR | 23-SEP-2002;   | 2002US-0412955P. |
| PR | 30-SEP-2002;   | 2002US-0415195P. |
| PR | 23-OCT-2002;   | 2002US-0420627P. |
| PR | 23-OCT-2002;   | 2002US-0420718P. |
| PR | 24-OCT-2002;   | 2002US-0420852P. |
| PR | 31-OCT-2002;   | 2002US-0422750P. |
| PR | 01-NOV-2002;   | 2002US-0423095P. |
| PR | 05-NOV-2002;   | 2002US-0423748P. |
| XX | (CURA-) CURAGEN CORP.  |                  |
| PI | Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;              |                  |
| PI | Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;                  |                  |
| PI | Ellerman K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;                  |                  |
| PI | Grosse WM, Gunther E, Guo X, Gusev VY, Herkmann JL, Ji W, Kekuda R;          |                  |
| PI | Khrantsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;             |                  |
| PI | MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;               |                  |
| PI | Milliet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;          |                  |
| PI | Rieger DK, Rothenberg ME, Shenoy SG, Shinkets RA, Smithson G;                |                  |
| PI | Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;          |                  |
| PI | Wolenc AR, Zhong M, Zhong H;   |                  |
| XX | WPI; 2004-053467/05.   |                  |
| DR | N-PSDB; ADH41616.  |                  |
| XX | New NOVX polypeptides and nucleic acid molecules useful for preventing or    |                  |
| PT | treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,             |                  |
| PT | atherosclerosis or diabetes, in chromosome mapping, tissue typing or in      |                  |
| PT | pharmacogenomics.  |                  |
| XX | Claim 2; SEQ ID NO 170; 1503pp; English.                                     |                  |
| PS | The invention relates to 566 new isolated human polypeptides and their       |                  |
| CC | encoding genes, sequences that are at least 95% identical to these or        |                  |
| CC | polypeptides comprising one or more conservative substitutions in these. The |                  |
| CC | polypeptide, polynucleotide and antibodies against the polypeptides are      |                  |
| CC | useful in diagnosing, treating or preventing NOVX-associated disorders,      |                  |
| CC | e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,         |                  |

CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,  
CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.  
CC The nucleic acids are further used as hybridization probes, in chromosome  
CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The  
CC polypeptides are also useful as vaccines. This sequence represents an  
CC example of the polypeptide of the invention.  
XX  
SQ

Sequence 740 AA;

Query Match 31.7%; Score 3804; DB 8; Length 740;

Best Local Similarity 99.7%; Pred. No. 8.4e-221;

Matches 734; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

|    |      |   |      |
|----|------|---|------|
| QY | 305  | PEGPPONCVTGNITGKSFILWDPPPIIVTKPSRYVELYGPSCRILDNSTKDLKFAFTNL   | 364  |
| Db | 3    | PEGPPONCVTGNITGKSFILWDPPPIIVTKPSRYVELYGPSCRILDNSTKDLKFAFTNL   | 62   |
| QY | 365  | TPFTMDVYIAAETSAGTGPKSNISVFTPPDVGAVFDLQAEVSTQVRIITWKKPRQPN     | 424  |
| Db | 63   | TPFTMDVYIAAETSAGTGPKSNISVFTPPDVGAVFDLQAEVSTQVRIITWKKPRQPN     | 122  |
| QY | 425  | GIINQYRVKVLVPETGIIILENTLLTGNNEYINDPMAPEIVNIPEPMVGLYEGSAEMSSDL | 484  |
| Db | 123  | GIINQYRVKVLVPETGIIILENTLLTGNNEYINDPMAPEIVNIPEPMVGLYEGSAEMSSDL | 182  |
| QY | 485  | HSLATFIYNSHPDKNPPARRAEDQTSVVVTRNQVITDIAEQLSYVIRRLVPFTEHMI     | 544  |
| Db | 183  | HSLATFIYNSHPDKNPPARRAEDQTSVVVTRNQVITDIAEQLSYVIRRLVPFTEHMI     | 242  |
| QY | 545  | SVSAFTIMGEGPTVLVSRTRQOVPSIKIINYKNISSSILLYWDPPEYPNGKITHYTI     | 604  |
| Db | 243  | SVSAFTIMGEGPTVLVSRTRQOVPSIKIINYKNISSSILLYWDPPEYPNGKITHYTI     | 302  |
| QY | 605  | YAMELDTNRAFOITIDNSFLITGLKKYTKYKRVAASTHGGSSLSSEENDIFVRTSEDE    | 664  |
| Db | 303  | YAMELDTNRAFOITIDNSFLITGLKKYTKYKRVAASTHGGSSLSSEENDIFVRTSEDE    | 362  |
| QY | 665  | PSSSQDVEIDVTADRIKWSPEKPGNGIIAYEVLYKNIDTLYKNKNTSTTDIILRNL      | 724  |
| Db | 363  | PSSSQDVEIDVTADRIKWSPEKPGNGIIAYEVLYKNIDTLYKNKNTSTTDIILRNL      | 422  |
| QY | 725  | RPHLYNISVRSYTRFGHGNQVSSLSVRTSETVPDSAPENITYKNISSGEIELSLFPPS    | 784  |
| Db | 423  | RPHLYNISVRSYTRFGHGNQVSSLSVRTSETVPDSAPENITYKNISSGEIELSLFPPS    | 482  |
| QY | 785  | SPNGIIKKYTYLKRNGNEERTINTSLTQNIKVLKXYTOYIIIEVSASTLKGEGVRSAP    | 844  |
| Db | 483  | SPNGIIKKYTYLKRNGNEERTINTSLTQNIKVLKXYTOYIIIEVSASTLKGEGVRSAP    | 542  |
| QY | 845  | ISILTEADAPDSPQDFSVKQLSGVTVKLSWQPPLEPNGIILYYTVVWNRSSLKTINVT    | 904  |
| Db | 543  | ISILTEADAPDSPQDFSVKQLSGVTVKLSWQPPLEPNGIILYYTVVWNRSSLKTINVT    | 602  |
| QY | 905  | ETSLEISLDLYNVEYSAVYTASTRFGDKTGSNIISFQTEGAPSDPPKDVYVYANLSSSS   | 964  |
| Db | 603  | ETSLEISLDLYNVEYSAVYTASTRFGDKTGSNIISFQTEGAPSDPPKDVYVYANLSSSS   | 662  |
| QY | 965  | IILFWTPPSKPGNGIIOYYSVYVNTSGTFMNFTHLTDNDNMTVSTIIDKLTIFSYI      | 1024 |
| Db | 663  | IILFWTPPSKPGNGIIOYYSVYVNTSGTFMNFTHLTDNDNMTVSTIIDKLTIFSYI      | 722  |
| QY | 1025 | TFWLTASTSVGNGNKS 1040   |      |
| Db | 723  | TFWLTASTSVGNGNKS 738  |      |

Search completed: June 30, 2005, 14:59:18

Job time : 249 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 14:54:09 ; Search time 58 Seconds  
(without alignments)  
2948.638 Million cell updates/sec

Title: US-10-673-885-2  
Perfect score: 11985  
Sequence: 1 MDLFIILLFLLFGTSETQVD.....MDSLAMEGDVLEWEETM 2291

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Issued Patents AA:\*
- 1: /cgn2\_6/prodata/1/iaa/5A COMB.pep:\*
  - 2: /cgn2\_6/prodata/1/iaa/5B COMB.pep:\*
  - 3: /cgn2\_6/prodata/1/iaa/6A COMB.pep:\*
  - 4: /cgn2\_6/prodata/1/iaa/6B COMB.pep:\*
  - 5: /cgn2\_6/prodata/1/iaa/PCTUS COMB.pep:\*
  - 6: /cgn2\_6/prodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description        |
|------------|--------|-------------|--------|----|--------------------|
| 1          | 11985  | 100.0       | 2291   | 4  | US-09-822-871-2    |
| 2          | 10536  | 87.9        | 2301   | 4  | US-09-822-871-4    |
| 3          | 4535.5 | 37.8        | 898    | 1  | US-08-036-210-22   |
| 4          | 4535.5 | 37.8        | 898    | 2  | US-08-449-609-22   |
| 5          | 4535.5 | 37.8        | 898    | 4  | US-09-361-096A-22  |
| 6          | 1903.5 | 15.9        | 402    | 1  | US-08-036-210-15   |
| 7          | 1903.5 | 15.9        | 402    | 2  | US-08-449-609-15   |
| 8          | 1902.5 | 15.9        | 401    | 4  | US-09-361-096A-15  |
| 9          | 1713.5 | 14.3        | 322    | 1  | US-08-036-210-11   |
| 10         | 1713.5 | 14.3        | 322    | 2  | US-08-449-609-11   |
| 11         | 1713.5 | 14.3        | 322    | 4  | US-09-361-096A-11  |
| 12         | 1538.5 | 12.8        | 289    | 4  | US-09-361-096A-47  |
| 13         | 1194.5 | 10.0        | 334    | 1  | US-08-036-210-7    |
| 14         | 1194.5 | 10.0        | 334    | 2  | US-08-449-609-7    |
| 15         | 1194.5 | 10.0        | 334    | 4  | US-09-361-096A-7   |
| 16         | 1154   | 9.6         | 1911   | 1  | US-08-800-825A-5   |
| 17         | 1154   | 9.6         | 1911   | 2  | US-08-449-609-9    |
| 18         | 1154   | 9.6         | 1911   | 3  | US-09-158-657-5    |
| 19         | 1154   | 9.6         | 1911   | 5  | PCT-US94-10166-5   |
| 20         | 1107   | 9.2         | 1997   | 4  | US-09-949-016-6275 |
| 21         | 1084.5 | 9.0         | 1903   | 4  | US-09-949-016-8049 |
| 22         | 942    | 7.9         | 176    | 1  | US-08-036-210-9    |
| 23         | 942    | 7.9         | 176    | 2  | US-08-449-609-9    |
| 24         | 942    | 7.9         | 176    | 4  | US-09-361-096A-9   |
| 25         | 836    | 7.0         | 1188   | 1  | US-08-201-697-4    |
| 26         | 834.5  | 7.0         | 1246   | 4  | US-09-949-016-8051 |
| 27         | 834.5  | 7.0         | 1246   | 4  | US-09-949-016-8052 |

|    |       |     |      |   |                    |                   |
|----|-------|-----|------|---|--------------------|-------------------|
| 28 | 834.5 | 7.0 | 1246 | 4 | US-09-949-016-8053 | Sequence 8053, Ap |
| 29 | 834.5 | 7.0 | 1246 | 4 | US-09-949-016-8054 | Sequence 8054, Ap |
| 30 | 834.5 | 7.0 | 1246 | 4 | US-09-949-016-8055 | Sequence 8055, Ap |
| 31 | 833   | 7.0 | 1501 | 2 | US-08-447-464-3    | Sequence 3, Appli |
| 32 | 833   | 7.0 | 1501 | 2 | US-08-716-679-3    | Sequence 3, Appli |
| 33 | 827.5 | 6.9 | 1187 | 1 | US-08-201-697-2    | Sequence 2, Appli |
| 34 | 823.5 | 6.9 | 1274 | 4 | US-09-949-016-8828 | Sequence 8828, Ap |
| 35 | 823.5 | 6.9 | 1274 | 4 | US-09-949-016-8829 | Sequence 8829, Ap |
| 36 | 823.5 | 6.9 | 1274 | 4 | US-09-949-016-8830 | Sequence 8830, Ap |
| 37 | 823.5 | 6.9 | 1274 | 4 | US-09-949-016-8831 | Sequence 8831, Ap |
| 38 | 823.5 | 6.9 | 1274 | 4 | US-09-949-016-8832 | Sequence 8832, Ap |
| 39 | 820   | 6.8 | 1337 | 3 | US-08-854-585-2    | Sequence 2, Appli |
| 40 | 820   | 6.8 | 1337 | 4 | US-09-447-533-2    | Sequence 2, Appli |
| 41 | 820   | 6.8 | 1337 | 5 | PCT-US95-05512-2   | Sequence 2, Appli |
| 42 | 746.5 | 6.2 | 1711 | 2 | US-08-342-930-2    | Sequence 2, Appli |
| 43 | 740.5 | 6.2 | 1452 | 2 | US-08-449-644-8    | Sequence 8, Appli |
| 44 | 740.5 | 6.2 | 1452 | 2 | US-08-087-244A-8   | Sequence 8, Appli |
| 45 | 740   | 6.2 | 506  | 4 | US-09-949-016-8833 | Sequence 8833, Ap |

ALIGNMENTS

RESULT 1  
US-09-822-871-2  
; Sequence 2, Application US/09822871  
; Patent No. 6723547  
; GENERAL INFORMATION:  
; APPLICANT: WEBSTER, Marion et al  
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,  
; FILE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CL001219  
; CURRENT APPLICATION NUMBER: US/09/822,871  
; CURRENT FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 2291  
; TYPE: PRT  
; ORGANISM: Human  
US-09-822-871-2

|                       |     |  |               |           |              |
|-----------------------|-----|--|---------------|-----------|--------------|
| Query Match           |     | 100.0%;  | Score 11985;  | DB 4;     | Length 2291; |
| Best Local Similarity |     | 100.0%;  | Pred. No. 0;  |           |              |
| Matches 2291;         |     | Conservative 0;  | Mismatches 0; | Indels 0; | Gaps 0;      |
| QY                    | 1   | MDFLIILLFLLFGTSETQVDVSNVVPNGRIISIVIKYKEVCPMMQTVYTVQVRSKPSLEVLTTN | 60            |           |              |
| DB                    | 1   | MDFLIILLFLLFGTSETQVDVSNVVPNGRIISIVIKYKEVCPMMQTVYTVQVRSKPSLEVLTTN | 60            |           |              |
| QY                    | 61  | FLAGERVGSAGILLSNWTPPNNGRIISIVIKYKEVCPMMQTVYTVQVRSKPSLEVLTTN      | 120           |           |              |
| DB                    | 61  | FLAGERVGSAGILLSNWTPPNNGRIISIVIKYKEVCPMMQTVYTVQVRSKPSLEVLTTN      | 120           |           |              |
| QY                    | 121 | LNPGTTYEIKVAENSAGIGVSDPFLQTASSAFKGVNLTVEAYNASAVKLIWLPQ           | 180           |           |              |
| DB                    | 121 | LNPGTTYEIKVAENSAGIGVSDPFLQTASSAFKGVNLTVEAYNASAVKLIWLPQ           | 180           |           |              |
| QY                    | 181 | PNKGTISPKISVKHARGSIIVKDVIRVEDILTGLKPECNENSESFLWSTASPSPLGRV       | 240           |           |              |
| DB                    | 181 | PNKGTISPKISVKHARGSIIVKDVIRVEDILTGLKPECNENSESFLWSTASPSPLGRV       | 240           |           |              |
| QY                    | 241 | TPPSRTTHSSSTLTONEISSVWKEPISFVWTHLRPYVTVLFEVSAATTEAGYIDSTIVRT     | 300           |           |              |
| DB                    | 241 | TPPSRTTHSSSTLTONEISSVWKEPISFVWTHLRPYVTVLFEVSAATTEAGYIDSTIVRT     | 300           |           |              |
| QY                    | 301 | PSVPEGPQNCVGNITGKSFISILWDPPTIVTGKFSVRVELYGPSGRILDNSTKDKFA        | 360           |           |              |
| DB                    | 301 | PSVPEGPQNCVGNITGKSFISILWDPPTIVTGKFSVRVELYGPSGRILDNSTKDKFA        | 360           |           |              |
| QY                    | 361 | FTNLPFTMYDVYIAAFTSAGTGPKNISVFTTTPVPGAVFDQLAEVESTQVITWKKP         | 420           |           |              |

Db 361 FTNLTPFTMYDVVYIAAETSAGTGPCKSNISVFTPPDVPVGAFFDLQLAEEVESTQVRIWTKKP 420  
Qy 421 ROPNGLIINQYRKVLVPETGIIILENTLLTGNEYINDPMAPELVNIIVPMVGLYEGSAEM 480  
Db 421 ROPNGLIINQYRKVLVPETGIIILENTLLTGNEYINDPMAPELVNIIVPMVGLYEGSAEM 480  
Qy 481 SSDLHSLATFIYNHSHDKNFPARNRAEDQTSVPTTRNQYITDIAAQSLSYVIRRLVPFT 540  
Db 481 SSDLHSLATFIYNHSHDKNFPARNRAEDQTSVPTTRNQYITDIAAQSLSYVIRRLVPFT 540  
Qy 541 EHMISVSAFTINGEGPPTVLSVRTQOVPSIIKIINYKNISSSILLYWDPPPEYNGKIT 600  
Db 541 EHMISVSAFTINGEGPPTVLSVRTQOVPSIIKIINYKNISSSILLYWDPPPEYNGKIT 600  
Qy 601 HTYIYAMELDTNRAFOITIDNSFLITGLKKYTKYKMRVAASTHDESSLSSENDIFVRT 660  
Db 601 HTYIYAMELDTNRAFOITIDNSFLITGLKKYTKYKMRVAASTHDESSLSSENDIFVRT 660  
Qy 661 SEDEPSSQDVEVDVTADEIRLKWSPPEKNGIIIAEVLVYKNIDTLYMKNTSTTDII 720  
Db 661 SEDEPSSQDVEVDVTADEIRLKWSPPEKNGIIIAEVLVYKNIDTLYMKNTSTTDII 720  
Qy 721 LRNLPHLTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGEIELSF 780  
Db 721 LRNLPHLTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGEIELSF 780  
Qy 781 LPPSSNGIIKKYIYILKESNGNEERTINTSLTONIKVKYQYIIIVSASTLKGEV 840  
Db 781 LPPSSNGIIKKYIYILKESNGNEERTINTSLTONIKVKYQYIIIVSASTLKGEV 840  
Qy 841 RGAPTSILTEADAPDPPODFSVKQSLGVTVKLSWQPLEPNGIILYYVYVWNRSSSLKT 900  
Db 841 RGAPTSILTEADAPDPPODFSVKQSLGVTVKLSWQPLEPNGIILYYVYVWNRSSSLKT 900  
Qy 901 INVTTLSLSLDLYNVVEYSAYVTASTRFGDGKTSNIIISFOTPEGAPSDPPKDVYANL 960  
Db 901 INVTTLSLSLDLYNVVEYSAYVTASTRFGDGKTSNIIISFOTPEGAPSDPPKDVYANL 960  
Qy 961 SSSSIIILFWTPSPKNGIIQYYSVYRNTSGTFMQNFTLHETNDPNMTVSTIIDKLTI 1020  
Db 961 SSSSIIILFWTPSPKNGIIQYYSVYRNTSGTFMQNFTLHETNDPNMTVSTIIDKLTI 1020  
Qy 1021 FSYTFLWLTASTSVGNKSSDIIIEVYTDQDIPGEFVGNLTYESISSTAINVSWPPAPQ 1080  
Db 1021 FSYTFLWLTASTSVGNKSSDIIIEVYTDQDIPGEFVGNLTYESISSTAINVSWPPAPQ 1080  
Qy 1081 NGLVFFVYVSLIILOQTPRHVRPPLVTVYERSIYFDNLEKYTDYILKITPSTEGFSDTYTAQ 1140  
Db 1081 NGLVFFVYVSLIILOQTPRHVRPPLVTVYERSIYFDNLEKYTDYILKITPSTEGFSDTYTAQ 1140  
Qy 1141 LYIKTEEDVPETSPINTFNLSSTSVLSWDPVPKPGCALISYDILTLOGPNEYSFITS 1200  
Db 1141 LYIKTEEDVPETSPINTFNLSSTSVLSWDPVPKPGCALISYDILTLOGPNEYSFITS 1200  
Qy 1201 DNVIIIEELSPFTLYSFFFAAARTRKGLGPSSLIFFVYTDSESVPLAPPQNLTINCTSDFW 1260  
Db 1201 DNVIIIEELSPFTLYSFFFAAARTRKGLGPSSLIFFVYTDSESVPLAPPQNLTINCTSDFW 1260  
Qy 1261 LKMWSPPLPGGIVKVSFKIHEHETDTIYKNIISGFKTEAKLVGLEPVSTYSIRVSAFTK 1320  
Db 1261 LKMWSPPLPGGIVKVSFKIHEHETDTIYKNIISGFKTEAKLVGLEPVSTYSIRVSAFTK 1320  
Qy 1321 VGNQGNFNVVKFTTQESVDPVQNNQCMATSWQSVLVKWDPPKANGIIITOYMTVERN 1380  
Db 1321 VGNQGNFNVVKFTTQESVDPVQNNQCMATSWQSVLVKWDPPKANGIIITOYMTVERN 1380  
Qy 1381 STKVSQDDHMYTFIKLANTSYVFKVRASTAGDESTCHVSTLPETVPSVPTNIAFSD 1440  
Db 1381 STKVSQDDHMYTFIKLANTSYVFKVRASTAGDESTCHVSTLPETVPSVPTNIAFSD 1440  
Qy 1441 VQSTSATLWIRPDTTILGYFQNYKIITQLRAQCKEWESECEVYQKIQYLYEAHLTEST 1500  
Db 1441 VQSTSATLWIRPDTTILGYFQNYKIITQLRAQCKEWESECEVYQKIQYLYEAHLTEST 1500

Qy 1501 VYGLKFKRWYRFQVAASTNAGYGNASNWISTKTLPGPDGPPENVHVWVATSPFSISISWS 1560  
Db 1501 VYGLKFKRWYRFQVAASTNAGYGNASNWISTKTLPGPDGPPENVHVWVATSPFSISISWS 1560  
Qy 1561 EPAVITGPTCYLIDVKSVNDDEFNISPIKSNENKTIIEIKDLEIFTRYSVWITAFGTNIS 1620  
Db 1561 EPAVITGPTCYLIDVKSVNDDEFNISPIKSNENKTIIEIKDLEIFTRYSVWITAFGTNIS 1620  
Qy 1621 AAYVEGKSAEMIVTTLESAPKDPNNMTFOKIDPEVTKFQTLFPLPSPONGNIQVYQAL 1680  
Db 1621 AAYVEGKSAEMIVTTLESAPKDPNNMTFOKIDPEVTKFQTLFPLPSPONGNIQVYQAL 1680  
Qy 1681 VYREDDPTAVQIHNLSIIQKNTFVIAMLEGLKGGHTYNI SVYVNSAGAGKVPWRITM 1740  
Db 1681 VYREDDPTAVQIHNLSIIQKNTFVIAMLEGLKGGHTYNI SVYVNSAGAGKVPWRITM 1740  
Qy 1741 DIKAPARPKTPTPIYDATGKLLVTSTTTIRMPICYSDDDHGPICKNVQVLTATETGAQHD 1800  
Db 1741 DIKAPARPKTPTPIYDATGKLLVTSTTTIRMPICYSDDDHGPICKNVQVLTATETGAQHD 1800  
Qy 1801 GNVTKWYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYIIIGADNACMIPGHNEDKIC 1860  
Db 1801 GNVTKWYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYIIIGADNACMIPGHNEDKIC 1860  
Qy 1861 NGPLPKKKOYLFPKFRATNIMGQFTDSYSDPVKTLGEGLSERTVEIILSVTLCLISILL 1920  
Db 1861 NGPLPKKKOYLFPKFRATNIMGQFTDSYSDPVKTLGEGLSERTVEIILSVTLCLISILL 1920  
Qy 1921 GTAIFAFARIROKQKEGGTYSQDAEIIIDTKLKLQDLITVADLELKDRLTRPISKKSPL 1980  
Db 1921 GTAIFAFARIROKQKEGGTYSQDAEIIIDTKLKLQDLITVADLELKDRLTRPISKKSPL 1980  
Qy 1981 QHVEBELCTNNLKFQBESELSKFLQDLSSTDDADLPWRAKNRPFNIPYNNNNRVKLIJA 2040  
Db 1981 QHVEBELCTNNLKFQBESELSKFLQDLSSTDDADLPWRAKNRPFNIPYNNNNRVKLIJA 2040  
Qy 2041 DASVPGSDVINASYISGVLCNPEFIATQGLPCTVGDVFWVWVETRAKTLVMLTQCFEKG 2100  
Db 2041 DASVPGSDVINASYISGVLCNPEFIATQGLPCTVGDVFWVWVETRAKTLVMLTQCFEKG 2100  
Qy 2101 RIRCHQYWPEDNKPVTVEGDIIVITKLMEDVQIDWTIRDLKIERHGDGMTVRCNFTAWPE 2160  
Db 2101 RIRCHQYWPEDNKPVTVEGDIIVITKLMEDVQIDWTIRDLKIERHGDGMTVRCNFTAWPE 2160  
Qy 2161 HGVPENSAPLIHFVKLVASRAHDTPMIVHCSAGVGRGTGFIALDHLTHOHINDHDFVDI 2220  
Db 2161 HGVPENSAPLIHFVKLVASRAHDTPMIVHCSAGVGRGTGFIALDHLTHOHINDHDFVDI 2220  
Qy 2221 YGLVABLRSERMCWQNLQAQYIFLHQCILDLLSNKGSNOPICFVNVYSALQKMDSLDAMEG 2280  
Db 2221 YGLVABLRSERMCWQNLQAQYIFLHQCILDLLSNKGSNOPICFVNVYSALQKMDSLDAMEG 2280  
Qy 2281 DVELEWEETTM 2291  
Db 2281 DVELEWEETTM 2291

## RESULT 2

US-09-822-871-4  
; Sequence 4, Application US/09822871  
; Patent No. 6723547  
; GENERAL INFORMATION:  
; APPLICANT: WEBSTER, Marion et al  
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CL001219  
; CURRENT APPLICATION NUMBER: US/09/822, 871  
; CURRENT FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4

```

; LENGTH: 2301
; TYPE: PRT
; ORGANISM: Rat
US-09-822-871-4

Query Match      87.9%; Score 10536; DB 4; Length 2301;
Best Local Similarity 85.9%; Pred. No. 0;
Matches 1977; Conservative 162; Mismatches 151; Indels 12; Gaps 4;

QY 1 MDPLIIFLLFISGTSQVDSNVVPGTRIDITISSIS-TTYSPTVRIIVTNVTKPGPP 59
DB 1 MDPHRSFLFLLIGTSQVDSVSSPDGTGVDITLSSVSATTYSSPVSRITLATNVTKPGPP 60
QY 60 VFLAGRVGSAGILLISWNTPPNPNBRIISYIVKYKEVCPWMQTVYQVRKSDSLVLIT 119
DB 61 VFLAGRVGSAGILLISWNTPPNPNBRIISYIVKYKEVCPWMQTVYTRAPAKPDSLEVLIT 120
QY 120 NLNPGTTYIKVAEENSAGIGVPSDFPLQTAESAPGVNVLTVAEYNASAVKLWYLP 179
DB 121 NLNPGTTYIKVAEENNAGIGVPSDFPLQTAESAPGVNVLTVAEALNYSANVLWYLP 180
QY 180 QNGKITTSFKISVKHARSIGVVKDVSIRVEDILITGKLPECNENSSESFLWSTASPSPTLGR 239
DB 181 QNGKITTSFKISVKHARSIGVVKDVSIRVEDILITGKLPECNENSSESFLWSTASPSPTLGR 240
QY 240 VTPPSRTHSSSTLTQNEISSVWKEPISFVWTHLRPTTYLFEVSAATTEAGYIDSTIVR 299
DB 241 VTPVTRTQSSSTAARSKISSVWKEPISFVWTHLRPTTYLFEVSAATTEAGYIDSTIVR 300
QY 300 TPESVPEGPPONCVTGNITGKFSILWDPTTIVTGKFSYRVLYGPGSRIILNDSKOLXP 359
DB 301 TPESVPEGPPONCIMGVNTGKAFSISWDPTTIVTGKFSYRVLYGPGSRIILNDSKOLXP 360
QY 360 AFTNLTPFTMYDVYIAETSAGTGPKSNISVTPPDVPGCAVFDLQLAEBVESQVRLTWCK 419
DB 361 AFTNLTPFTMYDVYIAETSAGTGPKSNISVTPPDVPGCAVFDLQLAEBVESQVRLTWCK 420
QY 420 PRQNGIINQYRVKVLVPGTGIILENTLLTGNNEYINDPMAPEINIVBPMVGLYEGSAB 479
DB 421 PRQNGIISOYRVKVSLETVGVLENTLLTQDESINPMSPRIIMNLVDPMGIFEGSGE 480
QY 480 MSSDLHSLATFYINSHDPKXFNAPNRAEDQTPSVVTRNQYITDIAEOLSVIRRLVPF 539
DB 481 MSSDLHSPASFYINSHDPNDFPASTRAEBSQSPVTVTRNQYITDIAEOLSVIRRLVPF 540
QY 540 TEHMSVSFTIMGEGPPTVLSVRTQVPSISKIINYKNISSSSILLYWDPEYVNGKI 599
DB 541 TEHMSVSFTIMGEGPPTVLSVRTQVPSISKIINYKNISSSSILLYWDPEYVNGKI 600
QY 600 THYTIYABELDNRAFIQTIDNSFLITGLKYYTKYKMRVAASVTHDGESSLSEENDIFVR 659
DB 601 THYTIYATELDNRAFOMTVDNSFLITGLKYYTKYKMRVAASVTHDGESSLSEENDIFVR 660
QY 660 TSEDEPESQDVEVDVTADEIRLKWSPPEKPNGIIIAEVLVYKXNIDFLYMNKSTTDDI 719
DB 661 TPEDEPESQDQVQVTVGSFSELRLKWSPPEKPNGIIIAEVLVYKXNIDFLYMNKSTTDDI 720
QY 720 ILNRLPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITKNISSGEIELS 779
DB 721 IISDLKPYTLNISHSYTRLGHGNQVSSLLSVRTSETVPSAPENITKNISSGEIELS 780
QY 780 FLPPSPNGIIKKYTYILKRSNGNEERTINTTSLQNIKVLKYYQYIIEVSASTLKGEG 839
DB 781 FLPPSPNGIIKKYTYILKRSNGNEERTINTTSLQNIKVLKYYQYIIEVSASTLKGEG 840
QY 840 VRSAPISILTEEDAPSPQDFSVKQSGVTVKLSWQPLEPNGLIYTVVYVNRSSLK 899
DB 841 IRRSRPISILTEEDAPSPQDFSVKQSGVTVKLSWQPLEPNGLIYTVVYVNRSSLK 900
QY 900 TINVTSTLESLDLDYNNYISAVVTASTRFGDGKTCNSIISQTPBEGAPSDPPKOVYAN 959
DB 901 AINATEASLVSLDLDYNNYISAVVTASTRFGDGKTCNSIISQTPBEGAPSDPPKOVYAN 960
960 LSSSSIIILFWTPPSKNGIIQYYSVYRYNTSCTFMONETLHELHETLNDPNNMTVSTIIDLKT 1019
961 LSSSSIIILFWTPPSKNGIIQYYSVYRYNTSCTFMONETLHELHETLNDPNNMTVSTIIDLKT 1020
1020 IFSYTYFWLTASTSVGNKSSDIIIEVYTDQDIPGPGVGNLTYSSTAIHVSVPVPAQ 1079
1021 IFSYTYFWLTASTSVGNKSSDIIIEVYTDQDIPGPGVGNLTYSSTAIHVSVPVPAQ 1080
1080 PNGLVYVSLTQOT-PRHVRPPLVTVYERSYFONLEKYTDYILKITPSTKSGSDTYT 1138
1081 PNGLVYVSLTQOT-PRHVRPPLVTVYERSYFONLEKYTDYILKITPSTKSGSDTYT 1140
1139 AQLYKITEDEVPETSPINTFNKLSSTSVLSWDPVVKPENGAIISVDLTQCPNENYSFI 1198
1141 TQLHIKITEDEVPETSPINTFNKLSSTSVLSWDPVVKPENGAIISVDLTQCPNENYSFI 1200
1199 TSDNVIILBELSPFLYSSFAAARTRKGLGPSIILFFYTDSEVPPLAPPONLTLINCTSD 1258
1201 TSGNHIVLBELSPFLYSSFAAARTRKGLGPSIILFFYTDSEVPPLAPPONLTLINCTSD 1260
1259 VMLKWSPLPGGIYKVSFKIHEHETDTIYKNIISGFKTEAKLVGLEPVSYSIRVSAP 1318
1261 VMLTWSPLPGGIYKVSFKIHEHETDTIYKNIISGFKTEAKLVGLEPVSYSIRVSAP 1320
1319 TKVGNQFQSVNVKTTQESVDPVQVQWQCMATSQSVLVKWDPPKPKANGIITQYVTVS 1378
1321 TKVGNQFQSVNVKTTQESVDPVQVQWQCMATSQSVLVKWDPPKPKANGIITQYVTVS 1380
1379 RNSTKVSQDQMTYTIKLLANTSYVFKVRASTVYFVRASTVYFVRASTVYFVRASTV 1438
1381 RNSTKVSQDQMTYTIKLLANTSYVFKVRASTVYFVRASTVYFVRASTVYFVRASTV 1440
1439 SDVQSTSATLWIRPDTILGPFQNYKIITQLRAQCKEWESECEVYKIQIYLAHLTE 1498
1441 SNVQSTSATLWIRPDTILGPFQNYKIITQLRAQCKEWESECEVYKIQIYLAHLTE 1500
1499 ETVYGLKFRWYRFOVAASVAGNAGNASTKTLPGPDPDPDPDPDPDPDPDPDPDPDP 1558
1501 ETVYGLKFRWYRFOVAASVAGNAGNASTKTLPGPDPDPDPDPDPDPDPDPDPDPDP 1560
1559 RSEPAVITGPTCYLLDKVSDNDENFIISPLKSNENKTEIKDLAETRYSVVITAFVGN 1618
1561 RSEPAVITGPTCYLLDKVSDNDENFIISPLKSNENKTEIKDLAETRYSVVITAFVGN 1620
1619 ISAAVEGKSSAEMIVTLESAPDPPNNMTFQKIPDEVTKEFQLTFLPSPQNGNIQVYQ 1678
1621 VSRATYDQKSSAEVITLESVPKDPNNMTFQKIPDEVTKEFQLTFLPSPQNGNIQVYQ 1680
1679 ALVYREDPPTAVQIHNLSIIQKNTFTVIAMLEGLKGGHTYNIISVAVNSAGAPKPMRI 1738
1681 ALVYREDPPTAVQIHNLSIIQKNTFTVIAMLEGLKGGHTYNIISVAVNSAGAPKPMRI 1740
1739 TMDIKAPARKPTPIYDATGKLVTSTIIRMPICYSDHGHPIKXVQVLAETGAQ 1798
1741 TMDIKAPARKPTPIYDATGKLVTSTIIRMPICYSDHGHPIKXVQVLAETGAQ 1800
1799 HDGNTWKYDAYFNKARPYFTNEGPNPCCKTKFSGNEEYIIGADNACMIPOGNEOK 1858
1801 HDGNTWKYDAYFNKARPYFTNEGPNPCCKTKFSGNEEYIIGADNACMIPOGNEOK 1860
1859 ICNGPLKPKQYLFKFRATNMGQFTSDSDYSDPVKTLGEGLSERTVEIISVTLCILSII 1918
1861 ICNGPLKPKQYLFKFRATNMGQFTSDSDYSDPVKTLGEGLSERTVEIISVTLCILSII 1920
1919 LLGTAFAPARIQKQEGGTYSPDASIIDTKLQDLITVADLELDERLTR----- 1972
1921 LLGTAFAPARIQKQEGGTYSPDASIIDTKLQDLITVADLELDERLTR----- 1980
1973 ---PLSKSKSFLOHVEELCTNNLKFQEBFSELPKFLQDLSSDADLPNNRKNRPNIKP 2029
1981 SIKPSKSKSFLOHVEELCTNNLKFQEBFSELPKFLQDLSSDADLPNNRKNRPNIKP 2040
2030 YNNNRVKLADASVPGSDYINASYISGLPCNPFATQGPLGTGDFWRVWVWETRAKT 2089
```

Db 2041 Y-NNNRVKLIADVSLPGSDYINASYVSGYLCNEFIATQGPLPGTVGDFWRMWTETRTKT 2099  
QY 2090 LVMLTQCFKGRIRCHOYWPEDNKPVTYVFGDIVITKLMEDVDQIDWTIRDLKIERHGDQMT 2149  
Db 2100 LVMLTQCFKGRIRCHOYWPEDNKPVTYVFGDIVITKLMEDIQIDWTIRDLKIERHGDQMT 2159  
QY 2150 VRQCNFTAPHEGVGPENSAPLHFVKLVRASRAHDTTPMIVHCSAGVGRGTGVFIADHLT 2209  
Db 2160 VRQCNFTGPEKGVPENTTPLHFVKLVRTSRAHDTTPMVHCSAGVGRGTGVFIADHLT 2219  
QY 2210 QHINHDFVDIYGLVAELSRMCMVQNLAQYIFLHQICILLNKGSGNQPICFVNYSL 2269  
Db 2220 QHINHDFVDIYGLVAELSRMCMVQNLAQYIFLHQICILLNKGSGHQPVCVNYSTL 2279  
QY 2270 QKMSLDAMEGDVELEWEETM 2291  
Db 2280 QKMSLDAMEGDVELEWEETM 2301

RESULT 3  
US-08-036-210-22  
; Sequence 22, Application US/08036210  
; Patent No. 5585233  
; GENERAL INFORMATION:  
; APPLICANT: Moller, Niels P.H.  
; APPLICANT: Moller, Karin B.  
; APPLICANT: Ullrich, Axel  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASE  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/036,210  
; FILING DATE: 23-MAR-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mierock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7683-025  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 898 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-036-210-22

Query Match 37.8%; Score 4535.5; DB 1; Length 898;  
Best Local Similarity 97.1%; Pred. No. 2.3e-269;  
Matches 861; Conservative 0; Mismatches 1; Indels 25; Gaps 2;  
QY 1429 VPSVPTNIAFSDVQSTSATLTWIRPDTILGYFNQYKITTQLRAQCKWESECEVEYQKI 1488  
Db 13 VPSVPTNIAFSDVQSTSATLTWIRPDTILGYFNQYKITTQLRAQCKWESECEVEYQKI 72  
QY 1489 QYLXEAHLTEETVYGLKFRWYRFQVAASNAGYGNASNIWSTKTLPGPPDGPENHVHV 1548

Db 73 QYLXEAHLTEETVYGLKFRWYRFQVAASNAGYGNASNIWSTKTLPGPPDGPENHVHV 132  
QY 1549 ATSPFSISISWSEPAVITGPTCYLIDVKSVDNDEFNIFSIKSEENKTIIEIKOLEIFTRY 1608  
Db 133 ATSPFSISISWSEPAVITGPTCYLIDVKSVDNDEFNIFSIKSEENKTIIEIKOLEIFTRY 192  
QY 1609 SVVITAFGTNISAAYVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKEQLTFLPPS 1668  
Db 193 SVVITAFGTNISAAYVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKEQLTFLPPS 252  
QY 1669 QPNNGIQQVQALVYREDDPTAVQIHNLISIIQKTNITFVIAMLEGLKGGHTYNI SVYAVNSA 1728  
Db 253 QPNNGIQQVQALVYREDDPTAVQIHNLISIIQKTNITFVIAMLEGLKGGHTYNI SVYAVNSA 312  
QY 1729 GAGPKVPMRITMDIKAPARPKTKPTPIYDATOKGLVSTTTIIRMPICIYSDHGPINKV 1788  
Db 313 GAGPKVPMRITMDIKAPARPKTKPTPIYDATOKGLVSTTTIIRMPICIYSDHGPINKV 372  
QY 1789 QVLATETGAQHDGNTWKYDAYFNKARPYFTNEGFPNPPCTEGKTFSGNBEIYIIGADN 1848  
Db 373 QVLVETGAQHDGNTWKYDAYFNKARPYFTNEGFPNPPCTEGKTFSGNBEIYIIGADN 432  
QY 1849 ACMIPGNEDEKICNGPLKPKKQYLFKFRATNIMGQFTSDYSDPVTKLGEGLSERTVEIIL 1908  
Db 433 ACMIPGNEDEKICNGPLKPKKQYLFKFRATNIMGQFTSDYSDPVTKLGEGLSERTVEIIL 492  
QY 1909 SVTLCILSIIILLGTAIFAPARIQKQEGGTYSPQDAEIIDTKLKDQITVADLELXDE 1968  
Db 493 SVTLCILSIIILLGTAIFAPARIQKQEGGTYSPQDAEIIDTKLKDQITVADLELXDE 552  
QY 1969 RLTR-----PISKSKFLQHVHEELCTNNNLKQFEFSELPKF 2004  
Db 553 RLTRYSSFFFRKEIFVQLLSYRKSIIPIKSKSFLQHVHEELCTNNNLKQFEFSELPKF 612  
QY 2005 LQDLSSTADALPWNRAKRNFPNIKPYNNNNRVKLIADASVPGSDYINASYISGYLCPNEF 2064  
Db 613 LQDLSSTADALPWNRAKRNFPNIKPY-NNNRVKLIADASVPGSDYINASYISGYLCPNEF 671  
QY 2065 IATQGPLPGTVGDFWRMWTETRAKTLVMLTQCFKGRIRCHOYWPEDNKPVTYVFGDIVIT 2124  
Db 672 IATQGPLPGTVGDFWRMWTETRAKTLVMLTQCFKGRIRCHOYWPEDNKPVTYVFGDIVIT 731  
QY 2125 KLMEVDQIDWTIRDLKIERHGDQMTVROCNTFTAPHEGVGPENSAPLHFVKLVRASRAHD 2184  
Db 732 KLMEVDQIDWTIRDLKIERHGDQMTVROCNTFTAPHEGVGPENSAPLHFVKLVRASRAHD 791  
QY 2185 TTPMIVHCSAGVGRGTGVFIADHLTQHINHDFVDIYGLVAELSRMCMVQNLAQYIPL 2244  
Db 792 TTPMIVHCSAGVGRGTGVFIADHLTQHINHDFVDIYGLVAELSRMCMVQNLAQYIPL 851  
QY 2245 HQICILLNKGSGNQPICFVNYSAQKMSLDAMEGDVELEWEETM 2291  
Db 852 HQICILLNKGSGNQPICFVNYSAQKMSLDAMEGDVELEWEETM 898

RESULT 4  
US-08-449-609-22  
; Sequence 22, Application US/08449609  
; Patent No. 5952212  
; GENERAL INFORMATION:  
; APPLICANT: Moller, Niels P.H.  
; APPLICANT: Moller, Karin B.  
; APPLICANT: Ullrich, Axel  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASE  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.



ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449,609  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/036,210  
FILING DATE: 23-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7683-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 898 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-449-609-22

Query Match 37.8%; Score 4535.5; DB 2; Length 898;  
Best Local Similarity 97.1%; Pred. No. 2.3e-269;  
Matches 861; Conservative 0; Mismatches 1; Indels 25; Gaps 2;  
QY 1429 VPSVPTNIAFSDVQSTSATLTWIRPDTILGYFQNYKITTQLRAQCKEWESECEVEYQKI 1488  
DB 13 VPSVPTNIAFSDVQSTSATLTWIRPDTILGYFQNYKITTQLRAQCKEWESECEVEYQKI 72  
QY 1489 QLYEAHLTEETVYGLKKFRWYRFOVAASTNAGYGNASNIWSTKTLPGPPDPPENHVHV 1548  
DB 73 QLYEAHLTEETVYGLKKFRWYRFOVAASTNAGYGNASNIWSTKTLPGPPDPPENHVHV 132  
QY 1549 ATSPFSISISWSEPAVITGPTCYLIDVKSVDNDEFNISPISKEENKTIIEIKDLEIFTRY 1608  
DB 133 ATSPFSISISWSEPAVITGPTCYLIDVKSVDNDEFNISPISKEENKTIIEIKDLEIFTRY 192  
QY 1609 SVVITAFNGISAAYVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKQLTFLPPS 1668  
DB 193 SVVITAFNGISAAYVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKQLTFLPPS 252  
QY 1669 QPNNGIQVQALVYREDDPTAVQIHNLSTIIQKNTFVIAMLGKGGHTYNIISVAVNSA 1728  
DB 253 QPNNGIQVQALVYREDDPTAVQIHNLSTIIQKNTFVIAMLGKGGHTYNIISVAVNSA 312  
QY 1729 GAGPKVPMRITWIDIKAPAPKTKPTPIYDATGKLLVTSITITIRMPICYYSDDHGPIKNV 1788  
DB 313 GAGPKVPMRITWIDIKAPAPKTKPTPIYDATGKLLVTSITITIRMPICYYSDDHGPIKNV 372  
QY 1789 QVLATBTGAHQDGNVTKVYDAVFNKARPYFTNEGPNPCTEGKTKFSGNEBIYIIGADN 1848  
DB 373 QVLATBTGAHQDGNVTKVYDAVFNKARPYFTNEGPNPCTEGKTKFSGNEBIYIIGADN 432  
QY 1849 ACMPGNEDKICNGPLKPKQVLYLKFERNIMQFTDSYSDPVKTLGSLSERTVEIIL 1908  
DB 433 ACMPGNEDKICNGPLKPKQVLYLKFERNIMQFTDSYSDPVKTLGSLSERTVEIIL 492  
QY 1909 SVTLCILSIIILGTAIFAPARIQKQEGTYSPODAEIIIDTKLKDQILITVADLELXDE 1968  
DB 493 SVTLCILSIIILGTAIFAPARIQKQEGTYSPODAEIIIDTKLKDQILITVADLELXDE 552  
QY 1969 RLTR-----PISKKSFLQHVHEELCTNNNLKQFESELPKF 2004  
DB 553 RLTRYSSFFPRKEIFVIQLLSYRKSIKPISKKSFLQHVHEELCTNNNLKQFESELPKF 612

QY 2005 LQDLSSTADLPPNRAKRPFPNPKYNNNNRVKLIADASVPGSDYINASYISGILCPNEF 2064  
DB 613 LQDLSSTADLPPNRAKRPFPNPKY-NNNRVKLIADASVPGSDYINASYISGILCPNEF 671  
QY 2065 IATQGPLGTVGDFWRMWETRAKTLVMTLQCFEKGRIRCHQYWPEDNKPVTVFGDIVIT 2124  
DB 672 IATQGPLGTVGDFWRMWETRAKTLVMTLQCFEKGRIRCHQYWPEDNKPVTVFGDIVIT 731  
QY 2125 KLMEVDQIDWTIRDLKIERHGDCTVRCNFTAWPEHGVPENSAPLIHFVKLVRAASHD 2184  
DB 732 KLMEVDQIDWTIRDLKIERHGDCTVRCNFTAWPEHGVPENSAPLIHFVKLVRAASHD 791  
QY 2185 TTPTMIVHCSAGVGRGTGVFIALDHLTQHIINDHDFVDIYGLVAELRSMCMQVQNLAYIFL 2244  
DB 792 TTPTMIVHCSAGVGRGTGVFIALDHLTQHIINDHDFVDIYGLVAELRSMCMQVQNLAYIFL 851  
QY 2245 HOCILDLISNKGNOPICFVNYSAQKMDSLDAMEGDVELEWEETM 2291  
DB 852 HOCILDLISNKGNOPICFVNYSAQKMDSLDAMEGDVELEWEETM 898

RESULT 5

US-09-361-096A-22  
Sequence 22, Application US/09361096A  
Patent No. 6492495  
GENERAL INFORMATION:  
APPLICANT: MOLLER, NIELS P.H.  
APPLICANT: MOLLER, KARIN B.  
APPLICANT: ULLRICH, AXEL  
TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE  
FILE REFERENCE: 038602/0686  
CURRENT APPLICATION NUMBER: US/09/361,096A  
CURRENT FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: 08/449,609  
PRIOR FILING DATE: 1995-05-24  
PRIOR APPLICATION NUMBER: 08/036,210  
PRIOR FILING DATE: 1995-03-23  
NUMBER OF SEQ ID NOS: 53  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 22  
LENGTH: 898  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: PTPS31-RD#2  
US-09-361-096A-22

Query Match 37.8%; Score 4535.5; DB 4; Length 898;  
Best Local Similarity 97.1%; Pred. No. 2.3e-269;  
Matches 861; Conservative 0; Mismatches 1; Indels 25; Gaps 2;  
QY 1429 VPSVPTNIAFSDVQSTSATLTWIRPDTILGYFQNYKITTQLRAQCKEWESECEVEYQKI 1488  
DB 13 VPSVPTNIAFSDVQSTSATLTWIRPDTILGYFQNYKITTQLRAQCKEWESECEVEYQKI 72  
QY 1489 QLYEAHLTEETVYGLKKFRWYRFOVAASTNAGYGNASNIWSTKTLPGPPDPPENHVHV 1548  
DB 73 QLYEAHLTEETVYGLKKFRWYRFOVAASTNAGYGNASNIWSTKTLPGPPDPPENHVHV 132  
QY 1549 ATSPFSISISWSEPAVITGPTCYLIDVKSVDNDEFNISPISKEENKTIIEIKDLEIFTRY 1608  
DB 133 ATSPFSISISWSEPAVITGPTCYLIDVKSVDNDEFNISPISKEENKTIIEIKDLEIFTRY 192  
QY 1609 SVVITAFNGISAAYVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKQLTFLPPS 1668  
DB 193 SVVITAFNGISAAYVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKQLTFLPPS 252  
QY 1669 QPNNGIQVQALVYREDDPTAVQIHNLSTIIQKNTFVIAMLGKGGHTYNIISVAVNSA 1728  
DB 253 QPNNGIQVQALVYREDDPTAVQIHNLSTIIQKNTFVIAMLGKGGHTYNIISVAVNSA 312  
QY 1729 GAGPKVPMRITWIDIKAPAPKTKPTPIYDATGKLLVTSITITIRMPICYYSDDHGPIKNV 1788

```
Db 313 GAGPKVPMRITMDIKAPRKTKPTBIYDATGKLLVTSITTIIRMEICYYSDHGGPIKV 372
QY 1789 QVLATETGAHQDGNVTWDAFNKARPFTNEGFPNCPCTEGKTKFSNGBEIYIIGADN 1848
Db 373 QVLVTETGAHQDGNVTWDAFNKARPFTNEGFPNCPCTEGKTKFSNGBEIYIIGADN 432
QY 1849 ACMIPOGNEKICNGPLKPKQYLFKFRATNIMQFTDSYSDPVKTLGEGLSERTVEIIL 1908
Db 433 ACMIPOGNEKICNGPLKPKQYLFKFRATNIMQFTDSYSDPVKTLGEGLSERTVEIIL 492
QY 1909 SVTLCILSIIILGTAIFAFARIRQKQEGTYSPODAEIIDTKLKDQITVADLELKD 1968
Db 493 SVTLCILSIIILGTAIFAFARIRQKQEGTYSPODAEIIDTKLKDQITVADLELKD 552
QY 1969 RLTR-----PISKSFLOHVEELCTNNLKFQEFSELPKF 2004
Db 553 RLTRYSSFFFRKEIFVIQLSYRKSIRPKSFLOHVEELCTNNLKFQEFSELPKF 612
QY 2005 LODLSSTDLADLPWNAKRNFPNKPNNNRVKLIADASVPGSDYINASYISGYLCPNEF 2064
Db 613 LODLSSTDLADLPWNAKRNFPNKPYPY-NNNRVKLIADASVPGSDYINASYISGYLCPNEF 671
QY 2065 IATQGLPCTVGDVFWRMVWETRAKTLVMLTQCFEKGRIHQYWPEDNKPVTVFGDIVIT 2124
Db 672 IATQGLPCTVGDVFWRMVWETRAKTLVMLTQCFEKGRIHQYWPEDNKPVTVFGDIVIT 731
QY 2125 KLMEVDQIDWTIRDLKIERHGDGMTVRCQNFATWPEHGVPENSAPLIHFVKLVASRAHD 2184
Db 732 KLMEVDQIDWTIRDLKIERHGDGMTVRCQNFATWPEHGVPENSAPLIHFVKLVASRAHD 791
QY 2185 TTPMIVHCSAGVGRGTGFALDHLTOHINDHDFVDIYGLVAELRSERCMQVONLAQYIFL 2244
Db 792 TTPMIVHCSAGVGRGTGFALDHLTOHINDHDFVDIYGLVAELRSERCMQVONLAQYIFL 851
QY 2245 HQCILDLLSNKGSNPICFVNYSAEQKMDSLDAMEGDVELEWEETM 2291
Db 852 HQCILDLLSNKGSNPICFVNYSAEQKMDSLDAMEGDVELEWEETM 898
```

## RESULT 6

```
US-08-036-210-15
; Sequence 15, Application US/08036210
; Patent No. 5585233
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/036,210
; FILING DATE: 23-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
```

```
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 402 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-036-210-15

Query Match 15.9%; Score 1903.5; DB 1; Length 402;
Best Local Similarity 93.5%; Pred. No. 1.2e-108;
Matches 362; Conservative 0; Mismatches 0; Indels 25; Gaps 2;

QY 1929 RIRQKQEGGTYSPODAEIIDTKLKDQITVADLELKDRLTR----- 1972
Db 17 RIRQKQEGGTYSPODAEIIDTKLKDQITVADLELKDRLTRYSSFFFRKEIFVIQL 76
QY 1973 -----PISKSFLOHVEELCTNNLKFQEFSELPKFLODLSSTDLADLPWNAKRN 2024
Db 77 LSYRKSIRPKSFLOHVEELCTNNLKFQEFSELPKFLODLSSTDLADLPWNAKRN 136
QY 2025 PNKPNNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGLPCTVGDVFWRMVW 2084
Db 137 PNKPYPY-NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGLPCTVGDVFWRMVW 195
QY 2085 TRAKTLVMLTQCFEKGRIHQYWPEDNKPVTVFGDIVITKLMEVDQIDWTIRDLKIERH 2144
Db 196 TRAKTLVMLTQCFEKGRIHQYWPEDNKPVTVFGDIVITKLMEVDQIDWTIRDLKIERH 255
QY 2145 GDMCTVRCQNFATWPEHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGF 2204
Db 256 GDMCTVRCQNFATWPEHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGF 315
QY 2205 LDHLTOHINDHDFVDIYGLVAELRSERCMQVONLAQYIFLHQCILDLLSNKGSNPICFV 2264
Db 316 LDHLTOHINDHDFVDIYGLVAELRSERCMQVONLAQYIFLHQCILDLLSNKGSNPICFV 375
QY 2265 NYSALQKMDSLDAMEGDVELEWEETM 2291
Db 376 NYSALQKMDSLDAMEGDVELEWEETM 402
```

## RESULT 7

```
US-08-449-609-15
; Sequence 15, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,609
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/036,210
; FILING DATE: 23-MAR-1993
```

ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7683-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 402 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-449-609-15

Query Match 15.9%; Score 1903.5; DB 2; Length 402;  
Best Local Similarity 93.5%; Pred. No. 1.2e-108;  
Matches 362; Conservative 0; Mismatches 0; Indels 25; Gaps 2;  
QY 1929 RIRQKQEGGTYSPODAEIIIDTKLKLQDLITVADLELKDERLTR----- 1972  
Db 17 RIRQKQEGGTYSPODAEIIIDTKLKLQDLITVADLELKDERLTRSSFFRRKEIFVIQL 76  
QY 1973 -----PISKKSFLQHVVELCTNNLKFQEFSELPKFLQDLSSTADLPNRAKNRP 2024  
Db 77 LSYRKSIIKPISSKKSFLQHVVELCTNNLKFQEFSELPKFLQDLSSTADLPNRAKNRP 136  
QY 2025 PNKPYNNNRVKKLIADASVPGSDYINASYISGLCPNEFIATQGPLTGVDFWRMWE 2084  
Db 137 PNKPY-NNNRVKKLIADASVPGSDYINASYISGLCPNEFIATQGPLTGVDFWRMWE 195  
QY 2085 TRAKTLVMTQCFEKGIRCHQYWPEDNKPVTVFQDIVITKLMEDVQIDWTIRDLKIERH 2144  
Db 196 TRAKTLVMTQCFEKGIRCHQYWPEDNKPVTVFQDIVITKLMEDVQIDWTIRDLKIERH 255  
QY 2145 GDCMTVRQCNFTAPHEGVSPENSAPLIHFVKLVRAASRAHDTTPMIVHCSAGVGRGVFTA 2204  
Db 256 GDCMTVRQCNFTAPHEGVSPENSAPLIHFVKLVRAASRAHDTTPMIVHCSAGVGRGVFTA 315  
QY 2205 LDHLTQHINDHDFVDIYGLVAELSRMCMQVONLAQYIFLHCILDLLSNKGSNPICFV 2264  
Db 316 LDHLTQHINDHDFVDIYGLVAELSRMCMQVONLAQYIFLHCILDLLSNKGSNPICFV 375  
QY 2265 NYSALQKMSLDAMEGDEVELEWETT 2291  
Db 376 NYSALQKMSLDAMEGDEVELEWETT 402

RESULT 8  
US-09-361-096A-15  
Sequence 15, Application US/09361096A  
Patent No. 6492495  
GENERAL INFORMATION:  
APPLICANT: MOLLER, NIELS P.H.  
APPLICANT: MOLLER, KARIN B.  
APPLICANT: ULLRICH, AXEL  
TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE  
FILE REFERENCE: 038602/0686  
CURRENT APPLICATION NUMBER: US/09/361,096A  
CURRENT FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: 08/449,609  
PRIOR FILING DATE: 1995-05-24  
PRIOR APPLICATION NUMBER: 08/036,210  
PRIOR FILING DATE: 1995-03-23  
NUMBER OF SEQ ID NOS: 53  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 15  
LENGTH: 401  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-361-096A-15

Query Match 15.9%; Score 1902.5; DB 4; Length 401;  
Best Local Similarity 92.8%; Pred. No. 1.4e-108;  
Matches 361; Conservative 2; Mismatches 1; Indels 25; Gaps 2;  
QY 1927 FARIQKQEGGTYSPODAEIIIDTKLKLQDLITVADLELKDERLTR----- 1972  
Db 14 YSEIRKQKQEGGTYSPODAEIIIDTKLKLQDLITVADLELKDERLTRSSFFRRKEIFVI 73  
QY 1973 -----PISKKSFLQHVVELCTNNLKFQEFSELPKFLQDLSSTADLPNRAKN 2022  
Db 74 QLLSYRKSIIKPISSKKSFLQHVVELCTNNLKFQEFSELPKFLQDLSSTADLPNRAKN 133  
QY 2023 RPNPKPYNNNRVKKLIADASVPGSDYINASYISGLCPNEFIATQGPLTGVDFWRM 2082  
Db 134 RPNPKPY-NNNRVKKLIADASVPGSDYINASYISGLCPNEFIATQGPLTGVDFWRM 192  
QY 2083 WETRAKTLVMTQCFEKGIRCHQYWPEDNKPVTVFQDIVITKLMEDVQIDWTIRDLKIE 2142  
Db 193 WETRAKTLVMTQCFEKGIRCHQYWPEDNKPVTVFQDIVITKLMEDVQIDWTIRDLKIE 252  
QY 2143 RHGDCMTVRQCNFTAPHEGVSPENSAPLIHFVKLVRAASRAHDTTPMIVHCSAGVGRGV 2202  
Db 253 RHGDCMTVRQCNFTAPHEGVSPENSAPLIHFVKLVRAASRAHDTTPMIVHCSAGVGRGV 312  
QY 2203 IALDHLTQHINDHDFVDIYGLVAELSRMCMQVONLAQYIFLHCILDLLSNKGSNPIC 2262  
Db 313 IALDHLTQHINDHDFVDIYGLVAELSRMCMQVONLAQYIFLHCILDLLSNKGSNPIC 372  
QY 2263 FVNSALQKMSLDAMEGDEVELEWETT 2291  
Db 373 FVNSALQKMSLDAMEGDEVELEWETT 401

RESULT 9  
US-08-036-210-11  
Sequence 11, Application US/08036210  
Patent No. 5585233  
GENERAL INFORMATION:  
APPLICANT: MOLLER, NIELS P.H.  
APPLICANT: MOLLER, KARIN B.  
APPLICANT: ULLRICH, AXEL  
TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
TITLE OF INVENTION: PHOSPHATASE  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSER: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/036,210  
FILING DATE: 23-MAR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7683-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 322 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

```
;
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-036-210-11

Query Match 14.3%; Score 1713.5; DB 1; Length 322;
Best Local Similarity 99.7%; Pred. No. 3.7e-97;
Matches 319; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1972 RPISKSFLOHVEELCTNNLKFOEFSELPKFLQDLSSTADLPWNRKNRPNIKPN 2031
Db 4 RPISKSFLOHVEELCTNNLKFOEFSELPKFLQDLSSTADLPWNRKNRPNIKPY- 62

QY 2032 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMVMWETRAKTIV 2091
Db 63 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMVMWETRAKTIV 122

QY 2092 MLTQCFEKGRIHQWPNEDNKPVTVFQDIVITKLMEDVQIDWTIRDLKIERHGDGMTVR 2151
Db 123 MLTQCFEKGRIHQWPNEDNKPVTVFQDIVITKLMEDVQIDWTIRDLKIERHGDGMTVR 182

QY 2152 QCNFTAMPHEGVGPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVFIALDHLTOH 2211
Db 183 QCNFTAMPHEGVGPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVFIALDHLTOH 242

QY 2212 INDDHFDVIYGLVAELSRMCMQVQLAQYIFLHQCILDLLSNKGSNQPCFVNYSALQK 2271
Db 243 INDDHFDVIYGLVAELSRMCMQVQLAQYIFLHQCILDLLSNKGSNQPCFVNYSALQK 302

QY 2272 MDSLDAMEGDVELEWETTM 2291
Db 303 MDSLDAMEGDVELEWETTM 322
```

```
RESULT 10
US-08-449-609-11
; Sequence 11, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,609
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/036,210
; FILING DATE: 23-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
```

```
;
; LENGTH: 322 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-449-609-11

Query Match 14.3%; Score 1713.5; DB 2; Length 322;
Best Local Similarity 99.7%; Pred. No. 3.7e-97;
Matches 319; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1972 RPISKSFLOHVEELCTNNLKFOEFSELPKFLQDLSSTADLPWNRKNRPNIKPN 2031
Db 4 RPISKSFLOHVEELCTNNLKFOEFSELPKFLQDLSSTADLPWNRKNRPNIKPY- 62

QY 2032 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMVMWETRAKTIV 2091
Db 63 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMVMWETRAKTIV 122

QY 2092 MLTQCFEKGRIHQWPNEDNKPVTVFQDIVITKLMEDVQIDWTIRDLKIERHGDGMTVR 2151
Db 123 MLTQCFEKGRIHQWPNEDNKPVTVFQDIVITKLMEDVQIDWTIRDLKIERHGDGMTVR 182

QY 2152 QCNFTAMPHEGVGPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVFIALDHLTOH 2211
Db 183 QCNFTAMPHEGVGPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVFIALDHLTOH 242

QY 2212 INDDHFDVIYGLVAELSRMCMQVQLAQYIFLHQCILDLLSNKGSNQPCFVNYSALQK 2271
Db 243 INDDHFDVIYGLVAELSRMCMQVQLAQYIFLHQCILDLLSNKGSNQPCFVNYSALQK 302

QY 2272 MDSLDAMEGDVELEWETTM 2291
Db 303 MDSLDAMEGDVELEWETTM 322
```

```
RESULT 11
US-09-361-096A-11
; Sequence 11, Application US/09361096A
; Patent No. 6492495
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0586
; CURRENT APPLICATION NUMBER: US/09/361,096A
; CURRENT FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PTP-S31D
US-09-361-096A-11

Query Match 14.3%; Score 1713.5; DB 4; Length 322;
Best Local Similarity 99.7%; Pred. No. 3.7e-97;
Matches 319; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1972 RPISKSFLOHVEELCTNNLKFOEFSELPKFLQDLSSTADLPWNRKNRPNIKPN 2031
Db 4 RPISKSFLOHVEELCTNNLKFOEFSELPKFLQDLSSTADLPWNRKNRPNIKPY- 62

QY 2032 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMVMWETRAKTIV 2091
Db 63 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMVMWETRAKTIV 122
```

|    |      |  |      |
|----|------|--|------|
| Qy | 2092 | MLTQCFKGIIRCHQYWPEDNKVTVFGDVIITKLMEVDQIDMTIRDLKIERHGDCMTVR   | 2151 |
| Db | 123  | MLTQCFKGIIRCHQYWPEDNKVTVFGDVIITKLMEVDQIDMTIRDLKIERHGDCMTVR   | 182  |
| Qy | 2152 | QCNFATAPEHGVPENSAPLTHFVKLVRASRAHDTTPMIVHCSAGVGSTGVFIALDHLTQH | 2211 |
| Db | 183  | QCNFATAPEHGVPENSAPLTHFVKLVRASRAHDTTPMIVHCSAGVGSTGVFIALDHLTQH | 242  |
| Qy | 2212 | INDHDFVDIYGLVAELRSRMCVQNLAQYIFLHCITDILLSNKGSNQPICFVNYSALOK   | 2271 |
| Db | 243  | INDHDFVDIYGLVAELRSRMCVQNLAQYIFLHCITDILLSNKGSNQPICFVNYSALOK   | 302  |
| Qy | 2272 | MSLIDAMEGDVELEWETTM  | 2291 |
| Db | 303  | MSLIDAMEGDVELEWETTM  | 322  |

| Query Match               | 12.8%; Score 1538.5; DB 4; Length 289;                                     |
|---------------------------|--|
| Best Local Similarity     | 99.7%; Pred. No. 1.7e-86;  |
| Matches 286; Conservative | 0; Mismatches 0; Indels 1; Gaps 1;   |
| QY                        | 1972 RPI SKSFLQHV BELCTNNLKPQEPSELPKFLQDLSSTADLPNNRAKRNPNPKPN 2031         |
| DB                        | 4 RPI SKSFLQHV BELCTNNLKPQEPSELPKFLQDLSSTADLPNNRAKRNPNPKPY- 62             |
| QY                        | 2032 NNNRVKLIADASVPGSDYINASVIGYLCNENPATAQGPLPGTVGDFWRMVWETRKTLLV 2091      |
| DB                        | 63 NNNRVKLIADASVPGSDYINASVIGYLCNENPATAQGPLPGTVGDFWRMVWETRKTLLV 122         |
| QY                        | 2092 MLTQCFEKGRIHQYWPEDONKPVTVGDIIVITKL MEDVOI DM TIRDLKTERIGDCMTVR 2151   |
| DB                        | 123 MLTQCFEKGRIHQYWPEDONKPVTVGDIIVITKL MEDVOI DM TIRDLKTERIGDCMTVR 182     |
| QY                        | 2152 QCNFTAWPHEGVSPENSAPLIHFVKLVRA SRAHD TTPMIVHCSAGVGRGVTFVIALDHLTQH 2211 |
| DB                        | 183 QCNFTAWPHEGVSPENSAPLIHFVKLVRA SRAHD TTPMIVHCSAGVGRGVTFVIALDHLTQH 242   |
| QY                        | 2212 INDHDPVDIYGLVAELRSERCMQVNLQAOYIFLHCQILDLLSNKGSN 2258                  |
| DB                        | 243 INDHDFVDIYGLVAELRSERCMQVNLQAOYIFLHCQILDLLSNKGSN 289                    |

```

; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Niels P.H.
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/036,210
; FILING DATE: 23-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-036-210-7

Query Match 10.0%; Score 1194.5; DB 1; Length 334;
Best Local Similarity 99.5%; Pred. No. 2.4e-65;
Matches 220; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1972 RPISKSFLOHVSELCNNLKFEQBFSELPKFLQDLSSDADLPNNRKNRPNNIKPYN 2031
Db 4 RPISKSFLOHVSELCNNLKFEQBFSELPKFLQDLSSDADLPNNRKNRPNNIKPY- 62
Qy 2032 NNNRVKLADASVPGSDYINASYISGYLCNPFATQGPLPGTVGDFWRMVWETRAKTLV 2091
Db 63 NNNRVKLADASVPGSDYINASYISGYLCNPFATQGPLPGTVGDFWRMVWETRAKTLV 122
Qy 2092 MLTQCFEKGIRCHQWYPEDNKNKVTTFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVR 2151
Db 123 MLTQCFEKGIRCHQWYPEDNKNKVTTFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVR 182
Qy 2152 QCNFTAPWPGVSPENSAPLIHFKVLVRASRAHDTTTPMIVHC 2192
Db 183 QCNFTAPWPGVSPENSAPLIHFKVLVRASRAHDTTTPMIVHC 223

RESULT 14
US-08-449-609-7
; Sequence 7, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Niels P.H.
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449,609  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/036,210  
FILING DATE: 23-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7683-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 334 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-449-609-7

Query Match 10.0%; Score 1194.5; DB 2; Length 334;  
Best Local Similarity 99.5%; Pred. No. 2.4e-65;  
Matches 220; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
  
QY 1972 RPISKKSFLQHVVEELCTNNLKFQBEFSELPKFLQDLSSTDADLPWNRKNRFPNPKPYN 2031  
Db 4 RPISKKSFLQHVVEELCTNNLKFQBEFSELPKFLQDLSSTDADLPWNRKNRFPNPKPY- 62  
  
QY 2032 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMVMWETRAKTIV 2091  
Db 63 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMVMWETRAKTIV 122  
  
QY 2092 MLTQCFEKGRIRCHOYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVR 2151  
Db 123 MLTQCFEKGRIRCHOYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVR 182  
  
QY 2152 QCNFTAWPEHGVGPENSAPLIHFVKLVASRAHDTPPMIVHC 2192  
Db 183 QCNFTAWPEHGVGPENSAPLIHFVKLVASRAHDTPPMIVHC 223

## RESULT 15

US-09-361-096A-7  
Sequence 7, Application US/09361096A  
Patent No. 6492495  
GENERAL INFORMATION:  
APPLICANT: MOLLER, NIELS P.H.  
APPLICANT: MOLLER, KARIN B.  
APPLICANT: ULLRICH, AXEL  
TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE  
FILE REFERENCE: 038602/0686  
CURRENT APPLICATION NUMBER: US/09/361,096A  
CURRENT FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: 08/449,609  
PRIOR FILING DATE: 1995-05-24  
PRIOR APPLICATION NUMBER: 08/036,210  
PRIOR FILING DATE: 1995-03-23  
NUMBER OF SEQ ID NOS: 53

SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 7  
LENGTH: 334  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: PTP-S31C  
US-09-361-096A-7  
  
Query Match 10.0%; Score 1194.5; DB 4; Length 334;  
Best Local Similarity 99.5%; Pred. No. 2.4e-65;  
Matches 220; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
  
QY 1972 RPISKKSFLQHVVEELCTNNLKFQBEFSELPKFLQDLSSTDADLPWNRKNRFPNPKPYN 2031  
Db 4 RPISKKSFLQHVVEELCTNNLKFQBEFSELPKFLQDLSSTDADLPWNRKNRFPNPKPY- 62  
  
QY 2032 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMVMWETRAKTIV 2091  
Db 63 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMVMWETRAKTIV 122  
  
QY 2092 MLTQCFEKGRIRCHOYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVR 2151  
Db 123 MLTQCFEKGRIRCHOYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVR 182  
  
QY 2152 QCNFTAWPEHGVGPENSAPLIHFVKLVASRAHDTPPMIVHC 2192  
Db 183 QCNFTAWPEHGVGPENSAPLIHFVKLVASRAHDTPPMIVHC 223

Search completed: June 30, 2005, 15:06:10  
Job time : 62 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 14:45:54 ; Search time 65 Seconds  
(without alignments)  
3391.271 Million cell updates/sec

Title: US-10-673-885-2

Perfect score: 11985  
Sequence: 1 MDPLIFLLFLLFGTSETQVD.....MDSLDMEGDVELEWEEETM 2291

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID       | Description        |
|------------|--------|-------------|--------|----------|--------------------|
| 1          | 10536  | 87.9        | 2302   | 2 T14328 | protein-tyrosine-p |
| 2          | 1319   | 11.0        | 2029   | 1 TDPLK  | protein-tyrosine-p |
| 3          | 1188   | 9.9         | 1912   | 2 A56178 | protein-tyrosine-p |
| 4          | 1184   | 9.9         | 1894   | 2 C54689 | protein-tyrosine-p |
| 5          | 1181   | 9.9         | 1898   | 2 S46216 | leukocyte antigen- |
| 6          | 1173   | 9.8         | 1691   | 1 D54689 | protein-tyrosine-p |
| 7          | 1155.5 | 9.6         | 1897   | 1 TDHULK | leukocyte antigen- |
| 8          | 1150   | 9.6         | 1615   | 2 B49502 | protein-tyrosine-p |
| 9          | 1150   | 9.6         | 1767   | 2 A49502 | protein-tyrosine-p |
| 10         | 1140   | 9.5         | 1907   | 2 S50893 | protein-tyrosine-p |
| 11         | 1127   | 9.4         | 1557   | 2 D41214 | protein-tyrosine-p |
| 12         | 1127   | 9.4         | 1630   | 2 C41214 | protein-tyrosine-p |
| 13         | 1107   | 9.2         | 1997   | 1 S12050 | protein-tyrosine-p |
| 14         | 1079   | 9.0         | 1863   | 2 S46217 | protein-tyrosine-p |
| 15         | 1033   | 8.6         | 2051   | 2 T30938 | receptor tyrosine  |
| 16         | 882.5  | 7.4         | 1290   | 2 A56493 | leucocyte common a |
| 17         | 863    | 7.2         | 1238   | 2 S68700 | HPTP beta-like tyr |
| 18         | 855    | 7.1         | 1367   | 2 T21913 | hypothetical prote |
| 19         | 841    | 7.0         | 1226   | 2 JC7503 | protein-tyrosine-p |
| 20         | 840    | 7.0         | 1499   | 2 I50212 | protein-tyrosine-p |
| 21         | 834    | 7.0         | 1188   | 1 A57064 | protein-tyrosine-p |
| 22         | 833    | 7.0         | 1501   | 2 I58148 | protein-tyrosine-p |
| 23         | 827.5  | 6.9         | 1187   | 1 A53661 | protein-tyrosine-p |
| 24         | 823    | 6.9         | 1262   | 1 B48758 | protein-tyrosine-p |
| 25         | 822    | 6.9         | 1216   | 2 S60613 | protein-tyrosine-p |
| 26         | 822    | 6.9         | 1337   | 1 I38670 | protein-tyrosine-p |
| 27         | 820    | 6.8         | 1496   | 1 A48758 | protein-tyrosine-p |
| 28         | 807.5  | 6.7         | 1585   | 2 T19121 | probable protein-t |
| 29         | 800    | 6.7         | 2222   | 2 T13924 | sdh protein - frui |

RESULT 1

T14328

protein-tyrosine-phosphatase (EC 3.1.3.48) receptor type, GMCl precursor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: T14328  
J;Wright, M.B.; Hugo, C.; Seifert, R.; Distech, C.M.; Bowen-Pope, D.F.  
J; Biol. Chem. 273: 23929-23937, 1998  
A;Title: Proliferating and migrating mesangial cells responding to injury express a nov  
A;Reference number: Z17986; MUID:98395110; PMID:9727007  
A;Accession: T14328  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-2302 <WRI>  
A;Cross-references: UNIPROT:O88488; EMBL:AF063249; NID:g3300095; PID:g3300096; PIDN:AAC  
A;Experimental source: strain Wistar  
C;Genetics:  
A;Note: PTPRQ

C;Keywords: phosphoric monoester hydrolase  
F;1-18/Domain: signal sequence #status predicted <SIG>  
F;19-2302/Product: protein-tyrosine phosphatase receptor type, GMCl #status predicted <

|                       |     |  |                 |            |              |
|-----------------------|-----|--|-----------------|------------|--------------|
| Query Match           |     | 87.9%;   | Score 10536;    | DB 2;      | Length 2302; |
| Best Local Similarity |     | 85.9%;   | Pred. No. 0;    |            |              |
| Matches 1977;         |     | Conservative 162;  | Mismatches 151; | Indels 12; | Gaps 4;      |
| QY                    | 1   | MDPLIFLLFLLFGTSETQVDVNVVPGTRYDITISSIS-TTYTSPVTRIVTPNWKPGPP   | 59              |            |              |
| DB                    | 2   | MDFHFSFLLLIGTSSQVDVSSFDGTGDTLSSVSATYSSPSVRTLATNWKPGPP        | 61              |            |              |
| QY                    | 60  | VFLAGRVGSAGILLISWNTPPNPNRIISYIVKYKEVCPWMTVYTVQVRKPDLSLELLT   | 119             |            |              |
| DB                    | 62  | VFLAGRVGSAGILLISWNTPPNPNRIISYIVKYKEVCPWMTVYTVQVRKPDLSLELLT   | 121             |            |              |
| QY                    | 120 | NLNPGETTIKVAENAGISGVFSDPFLQTAESAPGVNLTVEAYNASAVKLIWYLP       | 179             |            |              |
| DB                    | 122 | NLNPGETTIKVAENAGISGVFSDPFLQTAESAPGVNLTVEAYNASAVKLIWYLP       | 181             |            |              |
| QY                    | 180 | QPNQKITSPKISVGHARSIGIVKVDVIRVEDILTGLKPECNENSESFLNSTASPSPTLGR | 239             |            |              |
| DB                    | 182 | QPNQKITSPKISVGHARSIGIVKVDVIRVEDILTGLKPECNENSESFLNSTASPSPTLGR | 241             |            |              |
| QY                    | 240 | VTPPSRTTSSSLTQNEISSVWKEPISFVWTHLRPTTYLFEVSAATTEAGYIDSTIVR    | 299             |            |              |
| DB                    | 242 | VTPVTRTQSSSTAARSKISSVWKEPISFVWTHLRPTTYLFEVSAATTEAGYIDSTIVR   | 301             |            |              |
| QY                    | 300 | TPESVPEGPQNCVTGNITGKFSILWDPPTVTGKFSYRVLYGPGSRIIDNSTKDLKF     | 359             |            |              |
| DB                    | 302 | TPESVPEGPQNCVTGNITGKFSILWDPPTVTGKFSYRVLYGPGSRIIDNSTKDLRF     | 361             |            |              |
| QY                    | 360 | AFTNLTPFTMYDVYIAAETSAGTGPKNISVFTPPDPVPGAVFDLQLAEVSTQVITWK    | 419             |            |              |



Db 362 AFTHLTPFTMYDVYAAETSAGVGPKNLSVFTPPDPVPGAVFDLQIAEVEATEIRITWRK 421  
QY 420 PQPQNGIINOYRKVLVPTGTGIIENLTLLTGNXEYINDPMAPEIVNIPEPMVGLYEGSAE 479  
Db 422 PQPQNGIISQYRKVSVLSTGVVLENTLLTGQDESISNPNMSEIMMLVDPMLGFTYEGSGE 481  
QY 480 MSSDLHSLATFYNSHPDKNFAPARNRAEDQTSPPVVTTRNQYITDIAEQLSVYIRLVPF 539  
Db 482 MSSDLHSPASFYNLPHNDPFPASTRAEQSPVVTTRNQYMTDITAEQLSVYVRLVPF 541  
QY 540 THTMTSVSAPTINGEPPPTVLVSVTRQOVPSSIKIINYKNISSSSILLYWDPEYPNGKI 599  
Db 542 THTMTSVSAPTINGEPPPTVLVSVTRQOVPSSIQIINYKNISSSSILLYWDPEYPNGKI 601  
QY 600 THTYIAMELDTNRAFQITTIQNSFLITGLKYTKYKMEVAASTHGDGSSLSRENDIFVR 659  
Db 602 THTYIATELDTNRAFQMTIVONSFLITGLKYTRYKMRVAASTHGDGSSLSRENDIFVR 661  
QY 660 TSEDEPESPQDVEVDVTADEIRLKWSPPEKPNGIIIAEVLVYKNIDTLYMKNSTTDDI 719  
Db 662 TPEDEPESPQDVQVTGVSPSELRLKWSPPEKPNGIIIAEVLVYQNADTLFVKNTSTDDI 721  
QY 720 ILRNLRPHLYNISVRSTRFGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGEIELS 779  
Db 722 IISDLKPYTLNISRYSYTRLGHGNQSSLLSVRTSETVPDSAPENITYKNISSGEIELS 781  
QY 780 FLPPSSPNGIISKYTYLXRSNGNEERTINTSLQNIKVLYKYTCYIIEVSASTLKGRG 839  
Db 782 FLPPSPNGIISKYTYLXRSNSHEARTINTSLAQITGGLKKYTHYVIEVSASTLKGRG 841  
QY 840 VRSAPISILTEEDAPSPQDPSVKQLSGVTVKLSWQPPLEPNGLIYTVVYVNRSSLK 899  
Db 842 IISRPSILITEEDAPSPQDPSVKQLSGVTVKLSWQPPLEPNGLIYTVVYVNDKSLR 901  
QY 900 TINVTETSELDLDVNYYSAYVTASTRFGDKTGSNIISFOTPEGAPSDPKQVYYAN 959  
Db 902 AINATEASLVSLDLDVNDYVACVTASTRFGDGNARSSIINFRTPEGEPSDPPNDVHYVN 961  
QY 960 LSSSSIIILFWTPSPKNGIIOQYSVYRNTSGTFMGNFTLHELTPNDPMNTVSTIIDKLT 1019  
Db 962 LSSSSIIILFWTPSPKNGIIOQYSVYQNTSGTFVQNFLLQVTKESDNTVVSARIYRLA 1021  
QY 1020 IFSYTYFWLTASTSVGNKSSDIEVYTDQDIPPEGFVGNLTYESISSTAINVSWVPPAQ 1079  
Db 1022 IFSYTYFWLTASTSVGNKSSDIIHYVTDQDIPPEGVGNLTYESISSTAIHVSWEPPSQ 1081  
QY 1080 PNGLVYVYSLILQOT-PRHVRPPLVYERSYFDNLEKYTDYILKITPSTKGFSDTYT 1138  
Db 1082 PNGLVFYVYSLNLQOSPPRHMIPLVLYENSIDFDLLEKYTDYIFKITPSTKGFSETYT 1141  
QY 1139 AOLYKITEEDVPETSPINTEKNLSSTSVLLSWDPVPKNGALISYDLTLQGPENYSPT 1198  
Db 1142 TQLHIKTEEDVPETSPINTEKNLSSTSVLLSWDPVPKNGALGLVHLTLQGPANHFTV 1201  
QY 1199 TSDNYIILBELSPFTLYSFFAAARTKGLGPSILFFYTTDESVPPLAPPQNLTILNCTSD 1258  
Db 1202 TSGNHVILBELSPFTLYSFFAAARTKGLGPSILFFYTTDESAPLAPPQNLTILNYSDF 1261  
QY 1259 VMLKWSPLPGGIVKIVGFKIHEHETDTIYKNIISGFKTEAKVLGLEVPSTVYIRVSAP 1318  
Db 1262 VMLTWSPLPGGIVKIVGFKIHEHETDTVYKNIISGLQTDAKLEGLVPSTVYSVSAP 1321  
QY 1319 TKVGNQNSNVKFTQTSVPDVMQCMATSWQSVLVKWDPPPKKANGIITQYVMVTE 1378  
Db 1322 TKVGNQNSNVVEFTQTSVPVPAVRNIECVARDWQSVVRWDPPTKNGIILHYIMITVG 1381  
QY 1379 RNSTKVSPODMYTFIKLANTSYVFKVRASTVAGSGDSTCHVSLTPETVPSVPTNIAF 1438  
Db 1382 GNSTKVSPPDPTFTFKLPTNSYVFEVRASTVAGSGNESRCDISTLPETVPSAPTNVAF 1441  
QY 1439 SDVQSTSATLTWRPTILGYFQNYKITTLQRAQCKWESEBECVYQKIQVLYEAHLTE 1498  
Db 1442 SNVQSTSATLTWTKPTDITFGYFQNYKITTLQRAQCKREWEPECIEHQDKQVLYEANOTE 1501

QY 1499 ETVYGLKKFRWYRFOVAASTNAGYGNASNWISTKTLPGPDPDPENHVHVAISPFISIS 1558  
Db 1502 ETVHGLKKFRWYRFOVAASTNVGYSNASEWISTQTLPGPDPDPENHVHVAISPFGINIS 1561  
QY 1559 WSEPAVITGPTCYLIDVKSVDNDEPNISPIKNEENKTIIEIKDLEIFTRYSVVITAFQTN 1618  
Db 1562 WSEPAVITGPTCYLIDVKSVDNDEPNISPLKNEENKTIIEINNLEVFTRYSVVITAFVGN 1621  
QY 1619 ISAAVVEGSSAEMLVTTLESAPKDPNNTFKIPDEVTKEQLTFLPPSQNGNIQVQ 1678  
Db 1622 VSRAYTDGKSSAEVITTTLESVPKDPNNTFKIPDEVTKEQLTFLPPSQNGNIRVQ 1681  
QY 1679 ALVYREDDPTAVQIHNLSIIQKNTNFVIANLSEGLKGGHTYINISVYAVNSAGAGPKVPMRI 1738  
Db 1682 ALVYREDDPTAVQIHNFSIIQKNTDTSIANLSEGLKGGHTYINISVYAVNSAGAGPKVQMRI 1741  
QY 1739 TMDIKAPAPKTKPTPIVDATOKLIVTSTTIIRMPICYSDDHGPIKNOVQLATBTGAQ 1798  
Db 1742 TMDIKAPAPKSKPIRIRDATGKLIVTSTTIIRMPICYNDHGGPIRNOVQLVABTGAQ 1801  
QY 1799 HOGNVTKWDAYFNKARPYFTNEGPNPCTEGKTKFSGNEEYIIGADNACMIPGNEK 1858  
Db 1802 OQGNVTKWDAYFNKARPYFTNEGPNPCTEGKTKFSGNEEYIIGADNACMIPGNEK 1861  
QY 1859 ICNGPLKPKKQYLKFRATNIMGQFTDSYSDPVKTLGSELSERTVEIILSVTLCLISII 1918  
Db 1862 ICNGPLKPKKQYLKFRATNVGQFTSEYSDPIKTLGSELSERTVEIILSVTLCLISII 1921  
QY 1919 LIGTAIFAPARIRQKQEGGYSPQDAEIIDTKLKLQDLITVADLELKDRLTR----- 1972  
Db 1922 LIGTAIFAFVRIRQKQEGGYSPRAEIIDTKFKLDQILITVADLELKDRLTRLLSYRK 1981  
QY 1973 ---PISKSKFLQHVSELCTNNLKFQEESELPKFLQDLSSTDAFLPWRKNRPPNIKP 2029  
Db 1982 SIKPISKSKFLQHVSELCTNSNLKQEESESELPKFLQDLSSTDAFLPWRKNRPPNIKP 2041  
QY 2030 YNNNRVKLIADASVPGSDYINASYISGYLCNEFIATQGPLGTGDFPWRVWETRAKT 2089  
Db 2042 Y-NNNRVKLIADVSLPGSDYINASYISGYLCNEFIATQGPLGTGDFPWRVWETRTKT 2100  
QY 2090 LVMLTQCFEKGIRCHQYWPEDNKPVTVPGDIVITKLMEDVQIDWTIRDLKIERHGDCMT 2149  
Db 2101 LVMLTQCFEKGIRCHQYWPEDNKPVTVPGDIVITKLMEDIQIDWTIRDLKIERHGDCMT 2160  
QY 2150 VQOCNFTAPREHGVENSAPLIHFVKLVRAHADTTPMIVHCSAGVGRGTGFIALDHLT 2209  
Db 2161 VQOCNFTAPREHGVENITPLIHFVKLVRTSAHDTTPMIVHCSAGVGRGTGFIALDHLT 2220  
QY 2210 QHINDHDFVDIYGLVAELSERMCMVQNLAQYIFLHQICILDLNKGSGNQIPCFVNYVAL 2269  
Db 2221 QHINNDFVDIYGLVAELSERMCMVQNLAQYIFLHQICILDLNKGSGHQPVCVFNYSTL 2280  
QY 2270 QKMDSLDAMEGDEVELEWEETTM 2291  
Db 2281 QKMDSLDAMEGDEVELEWEETTM 2302

## RESULT 2

TDFFLK

protein-tyrosine-phosphatase (EC 3.1.3.48) DLAR precursor - fruit fly (*Drosophila melanogaster*)  
N:Alternate names: leukocyte antigen-related protein  
C:Species: *Drosophila melanogaster*  
C:Date: 14-Dec-1990 #sequence\_revision 02-May-1994 #text\_change 09-Jul-2004  
C:Accession: A36182  
R:Streuli, M.; Krueger, N.X.; Tsai, A.Y.M.; Saito, H.  
Proc. Natl. Acad. Sci. U.S.A. 86, 8698-8702, 1989  
A:Title: A family of receptor-linked protein tyrosine phosphatases in humans and *Drosophila*  
A:Reference number: A36182; MUID:90046860; PMID:2554325  
A:Accession: A36182  
A:Molecule type: mRNA  
A:Residues: 1-2029 <STR>  
A:Cross-references: UNIPROT:P16621; GB:M27700; NID:g157811; PID:AAA28668.1; PID:g157811





Db 630 WQPP----- 633  
QY 935 GSNISFQTEGAPDPKDVVYANLSSSSIIILFWTPSPKNGIIQYISYVYRNTSGTFM 994  
Db 634 -----PVEKQNGIIIEYSIKYTAVDG--- 654  
QY 995 QNFTLHETNDPDMVTSTIIDLKLTIPSYTFWLTASTSVGNKNGKSDIIEVTDQDIPR 1054  
Db 655 EDDKPEHILG-IPSDTTKLLLEQLEKWTETRTVTAHTDVGCPESLSVL-INTNEDVPS 712  
QY 1055 GFVGNLTYSISSTAINVSW---VPPAQNGLVYVYVLILOOTPRHVRPPLV----- 1104  
Db 713 GPPRKVEEAVNSTSVKWSRSPVKNQKQINGVQVHVVRMENGEPKQPMKQVMDLAD 772  
QY 1105 -----TYERSIYPDNLEKYTDYILKITPSTKGFSDTYTQLYIKTEBEDVPETSPIN 1157  
Db 773 AQWEPDPTTEHDIMIISGLQPEYSLSLTVTAYTTKGDGARSKPKL-VSTTGAVGPKRLV- 830  
QY 1158 TPKNLSTSVLLSWDPVPKPGNGAISYDYLTLQPNENYFITSNDVYIIEELSPTLYSF 1217  
Db 831 -INHTQMTALIQWHPVDTFGPLOGYRLKF----- 860  
QY 1218 FFAARTKGLGRSSILFFVTDESPLAPPQNLTLINCTSDFVWLKWSPLPGGIVKVYS 1277  
Db 861 -----GRKQMEPDTLEFSEK----- 877  
QY 1278 FKIHETDTIYKNIISGFKTEAKLVGLBFPVSTYSIRVSAPTKVGNQFNSNVKFTQOE 1337  
Db 878 -----DHFTATDIHGAS-----YVFLSARNKVGFGEM--VKEISIFE 915  
QY 1338 SVP-DVQNMCMATSWQSVLVKWDPPKA--NGIITQY-----MVTVERNSTKV 1384  
Db 916 EYPTGPPQLHSEGGTSTSVQSWQPPVLAERNGLITKTYLLYRDINILPEMQ---LI 972  
QY 1385 SPQDHMYTIKLLANTSVYFKVRASTAGSGBDES-TCHVSTLPEYTPSV-PTNIAFSDVQ 1442  
Db 973 VPADTTMTLTGLKPDYTDYDKVRAHSTKSGPVPSPVQRTLP--VDQFARNFHVKAVM 1030  
QY 1443 STSATLTWRPDTILGFYFONYKIITQLRAQCKEWESECEVYQKIYLYEAHLTEETVY 1502  
Db 1031 KTSVLLSWEP-----ENYNSAMPEFKIL-----YDDGKWEV-VDGRATQKLV 1073  
QY 1503 GLKFKRWYFQVAASNAG--YGNASNWISTKTLPGPPDGPENHVHATSPFSISWS 1560  
Db 1074 NLKPEKSYFVL---TNRNSAGGLQHRVTAKTAP-----DVLRTKP----- 1112  
QY 1561 EPAVITGTCYILDVKSVDNDEFNIFSKSRENTIEIKLEIFTRYSVVITATGNIS 1620  
Db 1113 -----AFI----- 1115  
QY 1621 AAYVEGKSAEMIVTLESAPKDPNNMTFQKIPDBVTKFQLTFLPPSPNGNIQVYQAL 1680  
Db 1116 -----GKTNLDGMTV--QLPEVPAN-----ENIKGYIIIVPLKSRK-----F 1154  
QY 1681 VYREDDPTAVQIHNL--SIQKNTNTFVIAMLEGLKGGHTYNTSVVAVNSAGAKVPKPMRI 1738  
Db 1155 IKPWESPDEMBELDELKESRKR-----RSIRYGREVELK- 1189  
QY 1739 TMDIKAPARKTKPTPIYDATGKLLVSTTIIRMPICYSDDHGPKVQVLAETGNAQ 1798  
Db 1190 -----PYIAHFDVLPTFTL-----GDD-----K 1209  
QY 1799 HDGNVTWKYDAVFNKARPYFTNEGPNPCTEGTKFSGNEBEIYIIGADNACWIPGNEDEK 1858  
Db 1210 HVG-----GFTN-----KQLSQGEVFFVLA---VMEHASK 1239  
QY 1859 ICGNPLPKKQYLFKFRATNIMGQFTSDSYSPVKTLG-----EGLSERTVEIIL 1908  
Db 1240 M-----YATSPYSDPVVSMDLDPQITDEESGLI-----WVV 1271  
QY 1909 SVTLCLISILGTAIFAFARIRQKQEGGTYSQDAEIID-----TKLKLQILITVADLE 1964  
Db 1272 GPVLAVVFIICIVIAILLYKRAESDSRKSSTPNKKEIPSHHPTDPPVELRL-----NQ 1327

QY 1965 LKDERLTRPISKKSFLQHVVELCTNNNLKAFQBEFSELPKFLQDLSTDLPMNRAKNRP 2024  
Db 1328 TPGMASHPPIPILELADHIERLKANDNLKFSQYESIDPG-QOFTWEHNSLVNKPKRY 1386  
QY 2025 PNIKYNNNRKLITADASVPGSDYNASYISGLYCPNFIATQGLPGTVGDFPWWME 2084  
Db 1387 ANVIAY-DHSRVLSSAIEGIPGSDYVNYIDYRKQNIAYIATQGSPLPETFGDFWRMI 1445  
QY 2085 TRAKTLVMLTQCFEKRIRCHOYWPEDNKPVTVPFGDIVITKLMEDVQI-DWTIRDLKIER 2143  
Db 1446 QRSATVVMVKLEERSRVKCDQWP--SRGTTHGLVQVT-LLDTVELATYCVRTALYK 1502  
QY 2144 HG--DQMTVRQCNFTAWPEHGVPEPNSAPLIHFVKLVRASTRAHDTTPMIVHCSAGVGTGV 2201  
Db 1503 NGSSSEKREVROFTAMPDHPGVEHPTPFLAFLRRVKTNCNPPDAGPMVHCSAGVGTGC 1562  
QY 2202 FTALDHLTGHINDHDFDIYGLVLAELSERCMQVQLAQYIFLHCQILDLNKGNSQPI 2261  
Db 1563 FVIDAMLERIKHEKTVDIYGHVTLMRAQNYMVQTDQYIFTHDALLEAVTCGNTEVPA 1622  
QY 2262 CFVNYTSALQKMDSLDAMEGDVLEWE 2287  
Db 1623 RNL-YAIQKLTQIETGENVTGHELE 1647

## RESULT 4

C54689

protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta, splice form B precursor  
N/Alternate names: MPTP delta type B/C  
N/Contains: protein tyrosine phosphatase, receptor type delta, splice form C  
C/Species: Mus musculus (house mouse)  
C/Date: 25-Apr-1995 #sequence\_revision 19-May-1995 #text\_change 09-Jul-2004  
C/Accession: C54689; B54689  
R/Mizuno, K.; Hasegawa, K.; Katagiri, T.; Ogimoto, M.; Ichikawa, T.; Yakura, H.  
Mol. Cell. Biol. 13, 5513-5523, 1993  
A/Title: MPTP delta, a putative murine homolog of HPTP delta, is expressed in specialized  
A/Reference number: A54689; MUID:93360986; PMID:8355697  
A/Accession: C54689

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-1894 &lt;Miz&gt;

A/Cross-references: UNIPROT:Q64487

A/Experimental source: brain; splice form B

A/Note: sequence inconsistent with nucleotide translation

A/Note: sequence extracted from NCBI backbone (NCBIN:137486, NCBI:137487)

A/Accession: B54689

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-352,'H',354-535,'S',537-601,1002-1894 &lt;M12&gt;

A/Experimental source: brain; splice form C

A/Note: sequence inconsistent with nucleotide translation

A/Note: sequence extracted from NCBI backbone (NCBIN:136527, NCBI:136530)

C/Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology;  
ogy

C/Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester by  
F:45-107/Domain: immunoglobulin homology <IMM1>  
F:245-299/Domain: immunoglobulin homology <IMM2>

F:317-399/Domain: fibronectin type III repeat homology &lt;FN3A&gt;

F:1278-1894/Domain: leukocyte common antigen cytosolic domain homology &lt;LAC&gt;

F:1652-1874/Domain: protein-tyrosine-phosphatase homology &lt;PTP2&gt;

F:1536/Active site: Cys (phosphocysteine intermediate) #status predicted

F:1542/Binding site: substrate phosphate (Arg) #status predicted

F:1826/Active site: Cys (phosphocysteine intermediate) #status predicted

F:1832/Binding site: substrate phosphate (Arg) #status predicted

## Query Match

Best Local Similarity 9.9%; Score 1184; DB 2; Length 1894;

Matches 490; Conservative 289; Mismatches 712; Indels 662; Gaps 80;

QY 263 KEPISFVTHLRPYTYTLFEVSAATEAGYIDSTIVRPESVPEGPPQNCVGTGNTGKSF 322  
Db 12 RRPLSLLLT-----PFLCACAETPPRF-----TRTPVD-----QTGVSQGVASFIC 52

|    |      |  |      |
|----|------|--|------|
| Qy | 323  | SILWDP--PTIVTGKFSVRV-----ELYGPGSRILDNSTKOLKFAFTNLTPF-----        | 367  |
| Db | 53   | QATGDPRPKIWNKGGKVSQNRFEVIEFDDGSGVL-----RIQPLRTPRDE               | 100  |
| Qy | 368  | TMVDVYIAAETASAGTGPKSNISVTPPDVPCA--VFDL--QLAEVESTOV-----          | 413  |
| Db | 101  | AIYEC-VASNNGVEISVSTRILVRBQOIPRGFFTIDMGPOLKVVERTRATATMLCAASGN     | 159  |
| Qy | 414  | --BITWKKPRQP-----NGIINQYRKVLVPETGIIILENTLLTGNNEYINDMPAEIV        | 464  |
| Db | 160  | POPEITWFKDFLPVDTSNNGRIKQLR-----SESIG-----ALQIE                   | 196  |
| Qy | 465  | NIVBPMVGLYEGSABMSDLHSLA---TFIYNSHPKDNFPARNRAEDQTSVPVTRNOYI       | 521  |
| Db | 197  | QSEESDQGYECVATNSAGTRYSPANALYVRVRPFRPSIPTNH-----EIMPGGSVNI        | 250  |
| Qy | 522  | TDIAAEQLSVYIRRLVPETHEMISVSAFTINGEGP--PTVLVSVKTRQO-----           | 567  |
| Db | 251  | TCVA-----VGSMPYVKKWMLGABDLTPDEDDPIGRNVLELNDVQCSANYTCVAMSTL       | 303  |
| Qy | 568  | -VPSIKIINYKNI-----SSSILLYWDPPEYPNGKXITHYTIYAMBLDNTRA             | 614  |
| Db | 304  | GVIEAIAQITVKALPKPGTPVVTESATSIITLWDSGN--PGPVSYIIL--QEKPKNSE       | 359  |
| Qy | 615  | FQITIID-----NSFLITGLKYYTKYKRWAASTHDGESSISEENDIFVRTSEDBEPSPQ      | 670  |
| Db | 360  | EPYKEIDGIATTRYSVAGLSYSDYEFVRVAVNNGRGPASE--PVLTOTSEQTSPSSAPR      | 417  |
| Qy | 671  | DVEVIDVTADRIELKWSPPKXPGNGIIIAEYLYKNIIDTLYMKNTSTTDI-----ILRNL     | 724  |
| Db | 418  | DVQARMLUSTTILVQWKEPEEPNQIOGYRVYTWMDPTQHVNNMKGNVADSQITTLGNL       | 477  |
| Qy | 725  | RPHTLYNISVRSYTRFSGHGNQVSLLSVRTSETVPDSAPENITYKNISSGEBLSFLPPS      | 784  |
| Db | 478  | VPQKTSYKVLAFTSIGDG-FLSSDIQVITQGVPGQPLNFKAEPESETSILLSWTPPR        | 535  |
| Qy | 785  | SPNGIIKYYIYLYKRSN--GNEERINTSLTONTIKVLKXYQYIIIEVSATLKEGVRSR       | 843  |
| Db | 536  | EDT--IASYELVDRDGOGEQRITIBFGYSYRLQGLKPNLSLYYFRLSATSPQGLGAST       | 593  |
| Qy | 844  | PISILTEDADPSPPQPSVKQLSGVTVKLSWQPPLEPNGIILYYTVVWNRSSLKNTINV       | 903  |
| Db | 594  | EISARTWQ-KPSAPPQDISCTSPSTSIILSVWQPP-----                         | 627  |
| Qy | 904  | TETSLSLDLDYNYEYSAYVTASTRFGDKTGTSNIIISFQTPEGAPDPKDVVYANLSS        | 963  |
| Db | 628  | -----  | 627  |
| Qy | 964  | SIILFWTPPPKXNGIIQYYSVYVYRNTSGTFMONTLHETLNDPDMTVSIIDKLTIIFY       | 1023 |
| Db | 628  | -----PVEKONGIITEYSLUKYAAVQG--EDYKPHIEIIGNSSD-TTKYLLEOLEKQWTE     | 676  |
| Qy | 1024 | YTFWLTASTSVGNGKSDILEVTDODIPGCFVGNLTYESISSTAINVSW---VPPAQP        | 1080 |
| Db | 677  | YRITVTAHTDVGPPESLSVL-IRTDDEVSPGPRKVEAVEAVNATAVKVSWRSPVPNKQH      | 735  |
| Qy | 1081 | NGLVYVYVSLILOOT--PRH--VRPLVYTERSYIFONLEKYTIYILKTIPTSEKGSFDT      | 1136 |
| Db | 736  | QOIRGYQVHYVWMEGEPKSAMLKDVMLADAQDMIISGLQPETSYSLTVTAITYTKGDGAR     | 795  |
| Qy | 1137 | YTAOLYIKTEEDVPETSPINTFNKLSSTSVLLSWDPPVKPNGAII SYDLTLQCPNENYS     | 1196 |
| Db | 796  | SKPKL-VSTTGSGPKPKPLV--INHTQNTALIQWHPPVDTFG-----LQYRLKFG          | 844  |
| Qy | 1197 | FITSDNYIILELSPFTLYSPFAARTRKGLGSPSSIIFYFTDESVPPLAPPQNLFLINCTS     | 1256 |
| Db | 845  | -----RKONEPLTTLFESEKE-----                                       | 860  |
| Qy | 1257 | DFVWLKWSPPSLPGGIVKVKYSGFKIHEHETDIITYYKNISGPKTEAKI/VGLEPVSITYSRVS | 1316 |
| Db | 861  | -----DHPTATDIIHKGAS-----YVFRLS                                   | 879  |

RESULT 5  
S46216  
leukocyt  
N;Altern



Db 962 VHL---TLGLPDDTYDIKVAHTSKGAGPLSPISQSRTP-MEQVFAKNRVAAMKT 1017  
QY 1554 SLSISWEP-----AVTGPCTVLIDVKSVDNDEFNIFISKNEENKTIIEIKLEIFTRY 1608  
Db 1018 SVLLSEVPSYKSAV---PFKILYNGQSVEDGHSMRKL-----IADLQNTPEY 1064  
QY 1609 SVVITAFNGISAAVVEGSSABMIVTTLESAP-----KOPPNMTFQKT--PDEVTK 1659  
Db 1065 SFVL--MNRGTSAGGQHLVSRITAPDLLPQKPLPASAFIDGRFSLSPQVDDPSLVRW 1122  
QY 1660 FOLTFLPPSPQNGNIQVQALVYREDPTAVQIHNLSIIQKNTVFVIAMLEGLKGGHTYN 1719  
Db 1123 FYIVVVPIDRVGNL-----LAPRWSTPELELDEL-----LEALEQGEERQ 1164  
QY 1720 ISVAVNSAGAPKVPWRITMDIKAPARKPTPTPIYDAGTKLLVTSITTITRMPICYYS 1779  
Db 1165 -----RRRRQAEELKPYVAAQVDELPTFTL----- 1191  
QY 1780 DGHGPIKNVQVLATETGAQDGNVTKWYDAYFNKARPFTNFGFPNPPCTEGTKFSGNE 1839  
Db 1192 ---GDKKN-----YRGFYNR----- 1203  
QY 1840 EIYIIGADNACMTPGNEDKICNGPLPKQY----LKFFRATNIMQFTDSYSDPVKTL 1895  
Db 1204 -----PLSPDLGYCFVLASKPEMDQKRYASSPYSDIIVQ 1240  
QY 1896 GEGLSERTVEIILSVT---LCILSIIILGTAPAFARIQKQKGGTYSPODAIITDKL 1952  
Db 1241 VTPAQOQEBPEMLWVTGPPVLAVIILIVIAILLFKRK-----THSPSSKDEQSIGL 1293  
QY 1953 KLDQLITVADLELKDERLT-----RPSKSKFLQVHEELCTNNLKFQEFSEL-- 2001  
Db 1294 K-DLSLAHSDDPVEMRLNYPQPMGDHPPIPTIDLIADNIERKANDGLKFSQYESIDP 1352  
QY 2002 --PKFLQDLSSTADLPWNRKAKRFPNPKYNNNRNKLADASVPGSDYINASYISGYLC 2060  
Db 1353 GQOFTWENSEV---NKPKNRYANVAY-DHSRVLLTSIDGVFSGDYINANYIDGYRK 1407  
QY 2061 PNEFIATQGLPFGTGDWRWVWETRAKTLMLTQCPEKGRIRCHQWYEDNKNPVTVFGD 2120  
Db 1408 QNAYIATQGLPFGTGDWRWVWETRAKTLMLTQCPEKGRIRCHQWYEDNKNPVTVFGD 2120  
QY 2121 IVITKLMEDVQI-DWTIRDLKIERHG--DCMTVROCNFTAPRHGVPENSAPLIHFVKLV 2177  
Db 1466 IQVT-LVDVTELATYTMRTFALHKSSEKREURQFQWAPDHGVPEYPTPLAFRRV 1524  
QY 2178 RASRAHDTPPMIVHCSAGVGRGTGFIALDHLTQHINDHDVFDIYGLVABLRSERCMQVN 2237  
Db 1525 KACNPLDAGPMVHCSAGVGRGTGFIVIDAMLERMKHKTVDIYGHVTCWRSQRYMVT 1584  
QY 2238 LAQYIFLHCIDLLSNKSGNQPCFVNTSALQKM-----DSLDAWEGDVEL 2284  
Db 1585 EDQYVFIHEALLE-AAMCGHTEVLARNLYAHIOKLQGVPPGSESVTAMELEFKL 1636

RESULT 6  
D54689  
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta, splice form D precursor  
N:Alternate names: MPTP delta type D  
N:Contains: protein tyrosine phosphatase, receptor type delta, splice form A  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: D54689; A54689  
R:Mizuno, K.; Hasegawa, K.; Katagiri, T.; Ogimoto, M.; Ichikawa, T.; Yakura, H.  
Mol. Cell. Biol. 13, 5513-5523, 1993  
A:Title: MPTP delta, a putative murine homolog of Hprt delta, is expressed in specialized  
A:Reference number: A54689; MUID:93360986; PMID:8355697  
A:Accession: D54689  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1691 <MI2>  
A:Cross-references: UNIPROT:Q64487

A:Experimental source: brain  
A>Note: sequence inconsistent with nucleotide translation  
A>Note: sequence extracted from NCBI backbone (NCBIN:137486, NCBIPI:136537)  
A:Accession: A54689  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-398,799-1691 <MI2>  
A:Experimental source: brain  
A>Note: sequence inconsistent with nucleotide translation  
A>Note: sequence extracted from NCBI backbone (NCBIN:136522, NCBIPI:136524)  
C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology;  
ogy  
C:Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester hvy  
F:42-95/Domain: immunoglobulin homology <IMM3>  
F:114-196/Domain: fibronectin type III repeat homology <FN3A>  
F:1075-1691/Domain: leukocyte common antigen cytosolic domain homology <LAC>  
F:1449-1671/Domain: protein-tyrosine-phosphatase homology <PTP>  
F:1333/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:1339/Binding site: substrate phosphate (Arg) #status predicted  
F:1623/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:1629/Binding site: substrate phosphate (Arg) #status predicted

Query Match 9.8%; Score 1173; DB 1; Length 1691;  
Best Local Similarity 26.4%; Pred. No. 5e-50;  
Matches 426; Conservative 266; Mismatches 600; Indels 319; Gaps 72;

QY 781 LPPSS-----PGLIKKTYI-----YLRSGNEERT-----INTTSLTON-IKVLKK 822  
Db 32 IPTTHNEIMPSSVNIITCVAGSPMPYKVMLGAEGLTTPEDDMPIGRNVLELNDVRQSAN 91  
QY 823 YTYQIIEVSASTLKGEGVRSAPISILTEDADPSPPQDFSVKQLSGVTVKLSQWPPLEPN 882  
Db 92 YTC-----VAMSTL---GVIEA-IAQITVKALP-KPGTPTVTESTATISLTWDSG-NP- 140  
QY 883 GILYTYVYVMNRSS---LKTIN-VTETSLESLDLYNVEXSAYVTASTRFGDGKTSNI 938  
Db 141 GPVSYVIIQHKPKNEEPYKEIDGIATTRYSVAGLSYSDYDFRVVAVNNIGRGPASEPV 200  
QY 939 ISFOTPEGAPSPDPKDYVYANLSSSIIILFWTPPKPNGIIQYYSYTYENTSGTFMQNET 998  
Db 201 LT-QTSEQTPSSAPRDVQARMLSSITLVQWKEPEEPNGIQGYRYVYTMDFPTQHVNNWM 259  
QY 999 LHELTYNDFNMTVSTIIDKLTIFSYVFWLTASTSVGNGKSDIIEVYTDQIDPEGFVG 1058  
Db 260 KENVA-DSQITIGNLVPKT---YSVKVLAFTSIGDGLSSD-IQVITQVGP-GQPL 312  
QY 1059 NUTYSISSTAINVSWPPAQNGLVFYVSVILIQOTPRHVPPLVYR--RSIYFDNLE 1116  
Db 313 NPKAPESESTILLSWTPP-RSDTIASVE---LVDRDGDQGBEQRITIEFGTSYRLOGLK 368  
QY 1117 KYTDYILKITPTEKGFSDTYTAQLYIKT-----EEDVPETSPILNTFKMLSSTSVLL 1169  
Db 369 PNLSYFFRLSATSPQGLGAS-TAETISARTMQKPSAPPQDISCTSP-----SSTSILV 419  
QY 1170 SWDPP--VKPNGAIIISYDLTLQCPNENYSFISDNY-----IILEELSPFT 1213  
Db 420 SWQPPVPEKQNGIITEYSL-----KYAANDGEDYKPHIEIGNSDDTKYLLQLEKWT 472  
QY 1214 LYSFFAAARTRKGLGPFSSI-LFFYTDSESVPLAPPQNLTL--INCTSDFVWLKWSPPSLP- 1269  
Db 473 EYRITVTAHTDVGWPESLSVLIRTDDEDVPSGPPRKRVEAVNATA--VKVSWR-SPVEN 529  
QY 1270 ---GGI--VKVYSPKIHETDTIYKNSIGPKTEAKLV-GLEPSTYSIRVSAPTKVN 1323  
Db 530 KOHQGIRGQVHYVKNENGEPKSAAMLKOVMLADAQDMIIISGLQFETSYSLTVAITYTKGD 569  
QY 1324 GNQFSNVVAKFTQESVPD----VVQNMOCMATSWOSVLVKWPPPKKANGIITQYMYTVBR 1379  
Db 590 GAR-SKPKLVSTGTVGPKRVLVNH-----TQMNTALIOMHPVDVTFG-LQGYRLKFR 642  
QY 1380 -----NSTKVSFQDHMYTFIKLLANTSVYFKVRASTSAGEDESTCHVSTLPTETVPS-V 1432  
Db 643 KDMELPTLLEFSEKEDHFTATDIHKGASYVFRLSARNKVGFGSEVMVKIS-VPEEPTGFG 701





Db 376 AFRVLAVNSIGRPPSEAVRAKTEQAPSPRRVQARMLASTMLVQWPEEPENGLVR 435  
QY 887 YTYVY-----VWNRSSLKINVTETSL--ELSDLDYXNVEYSAYTASTFRGDGKT 934  
Db 436 GYRVYVTPDSRRPPNAWKHN-----TDAGLLTTVGSLPGITVSLRVLAFTAVGDGPP 489  
QY 935 GNIISFQTEGAPSPKDVVYANLSSSS--ILFWTPPSKNGIIOYQYVYVYRNTGTF 993  
Db 490 -SPTIQVKTOQGVAPAD--FQAEVESDTRIQLSWLLPPQER--ILMYELVYM-----AA 540  
QY 994 MONFTLHETNPDNMTVSTIIDKLTFISYTFWLTASTVGNKNSDIIIEYTDQDIP 1053  
Db 541 EDEDQKHVTFD---PTSSYTTLEDLPDLTYRFLAARSDMGV-VFTPTIEARTAQSTP 596  
QY 1054 EGFVGNLTYESSTAINSVWPPA--QPNGLVFFY---VSLILQOTPRHRVPLVATYE 1107  
Db 597 SAPPQKVMCSGISTVRVSWVPPPPADSRNGVITQYVAHEAVDGEDRGRHVVDGLSRH 656  
QY 1108 RSIYFDNLEKYTDYILKITPSTEKGFSDDTYTAQLYIKTEEDVPETSPITNTFNLSSTSV 1167  
Db 657 SSWDLVGLKEWTEYRVVWRAHTDVG--PGPESSPVLVRTDEDVSPGPRKVEPELNSTAV 715  
QY 1168 LLSWDDPV--KNGAIIISYDLTLQGNENYSFITSDNYIILBELSPFTLYSPFAAARTK 1225  
Db 716 HVYWKLPVPSKQHGIRGQVT-----YVRLNGEP-----R 747  
QY 1226 GLGSSILFFYTDSEVPLAPPQNLTLNCTDFVWLKWSPLPGGIVKVSFKIHEHT 1285  
Db 748 GL-----PII-----QDVLAAQ-----WRP-----ESEE 768  
QY 1286 DRIYKNISGFKTEAKLVGLEPVSTYSIRVSATFKVGNQPSNVVKTFTQESVPDVQON 1345  
Db 769 D-----YETISGLTPTYSVTVAATTKGDGAR--SKPKIVTTTGAQVG--RPT 815  
QY 1346 MQMATSWOSLVKWDPPKANGIITQYMWTVR-----NSTKVPQDHMTYFKILAN 1399  
Db 816 MMISTAMNTALLQWHPKELPCELLGYRLQYCRADARPNTIDFGKDDQHFVTGLHGK 875  
QY 1400 TSVFKVRASTAGDESTCHVSTLPETVPS--VPTNIAPSDVQSPSATLTWRP----- 1453  
Db 876 TTYIFLAANRAGLGEKEKEIRT--PEDLPSPQNLHVHTGLTSTTTELANDPPVLAER 934  
QY 1454 -DTILGYFQNYK-ITTLQRAQCKEWESECEVEYQKIQVLYEAHLTEETVYGLKPRWTR 1511  
Db 935 NGRILSYTVFRDINSQ-----QELQNIITD--TRFTLTGLKPDITYD 975  
QY 1512 FOVAASTNAGYNASNMISTKTLPGPPDGPENNVHVATSPSISISWSEP-----AVIT 1566  
Db 976 IKVRAWTSKSGPLSPSIQSRTMP--VEQVFAKNFRVAAAAMKTSVLLSWEPVDSYKSAV-- 1032  
QY 1567 GPTCYLIDVKSVDNDEFNISFKSNEENKIEIKLEIFTRYSVVITATGNIISAAYVEG 1626  
Db 1033 -PFKILYNGQSVEVDGHSRKL-----IADLPNTSEYFVLNM-----RG 1071  
QY 1627 KSSAEM-IVTTLESAPKDPN-----NMTFOKIPDEVTKFOLTFLPPSQ 1669  
Db 1072 SSAGGLQHLVSRITADLLPKPLPASAVIEDGRPLDSMPHQDPSLVRFVIVVPIR 1131  
QY 1670 PNGNIQVQALVYREDPTAVQIHNLISIIQKNTFVIMLEGLKGHTYINISVYAVNSAG 1729  
Db 1132 VGGSM-----LTPRWSTPEELEDL-----LEAIEQG----- 1159  
QY 1730 AGPKVPMRITMDIKAPARKTKETPIYDATGKLLVTSTITITRMPICYSDDHGPIKNVQ 1789  
Db 1160 -GEEQRRR-----RRQAERLKPVA--AQDLVLPETFTL-----GDKKN-- 1195  
QY 1790 VLATETGAQHDGNVTKWYDAYFNKARPYPFTNEGFPNPPCTEGKTKPSGNEEYIIGADNA 1849  
Db 1196 -----YGFYNR----- 1202  
QY 1850 CMIPGNEDKICNGLPKPKQY-----LFKFRATNMQFTDSYSDPVKTLGEBLSRTVE 1905  
Db 1203 -----FLSPDLISYQCFLASLKEPMDQKRYASSPYSDBIYVQVTPAQOQEBP 1249

QY 1906 IILSVT---LCILSILLGTATFAFARIQKQKEGGTYSPODAEIIDTKLKLDQLITVAD 1962  
Db 1250 EMLWTPGVLAVILIIIVIAILLFKRKA-----THSPSKDEQSIGLK--DSLAAHS 1301  
QY 1963 LELKDBRLT-----RPISKKFLQHVVELCTNNNLKFOBESELSLPKFLODLSSTA 2013  
Db 1302 DPVEMRRLNYQTPGMRDHPPIITDLADNIELRKANDGLKFQSEYESIDPG--QQFTWNS 1360  
QY 2014 DLPWNRKRNFRPNKPYNNNRVKLIADASVPSGVINASYISYGLCPNEFIATQCPPLG 2073  
Db 1361 NLEVNKPKRYANVIAY--DHSRVILITSIDGVPSTYINANYIDGKVKQWAYIATQGPLPE 1419  
QY 2074 TVGDFWRMWWEVTRAKTLVMLTQCFFKGRIRCHQYQVPEDNKPVTVFGDIVITKLMEVQI- 2132  
Db 1420 TWGDFWRMWWEQRTATVVMVTRLEKSRVKCQYWPA--RGTTETCGLIQVT--LLDITVELA 1476  
QY 2133 DWTIRDLKIERHG--DCMTVROCNFTAMPEHGVENSAPLIHIFVKLVASRAHDTTPMIV 2190  
Db 1477 TVTVRTFALHKSSEKRELQFQMAWPDHGVPEYPTPILAFRRVKACNPLDAGPMV 1536  
QY 2191 HCSAGVGTGVFIADHLTOHINDHDFVDIYGLVAELSERMCWQNLAOYIFLHOCIID 2250  
Db 1537 HCSAGVGTGCFIVIDAMLERMKHEKTVDIYGHVTCMSQRNMQVQTEQYVFIHEALLE 1596  
QY 2251 LLSNKGSNQPICFVNVYSALQKM-----DSLDAWEGDVEL 2284  
Db 1597 ATNCGHTEVPARNL-YAHIQKLGQVPPGESVTAMELEPKL 1635

## RESULT 8

B49502

protein-tyrosine-phosphatase (EC 3.1.3.48) receptor type 4E, splice form B precursor -  
C:Species: Drosophila melanogaster  
C:Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
C:Accession: B49502  
J. Biol. Chem. 268, 23964-23971, 1993  
A:Title: Alternative splicing in a novel tyrosine phosphatase gene (DTP4E) of Drosophila.  
A:Reference number: A49502; MUID:94043220; PMID:8226938  
A:Accession: B49502  
A:Molecule type: mRNA  
A:Residues: 1-1615 <OON>  
A:Cross-references: UNIPROT:Q9W4F5; GB:120894  
C:Genetics:

A:Gene: FlyBase:Ptp4E  
A:Cross-references: FlyBase:FBgn0004368  
A:Introns: 1605/3

C:Superfamily: protein-tyrosine-phosphatase, receptor type 4E; fibronectin type III repeat  
C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; receptor  
F1254-1270/Domain: transmembrane #status predicted <TM>  
F1271-1615/Domain: intracellular #status predicted <INT>  
F1353-1573/Domain: protein-tyrosine-phosphatase homology <PTPI>  
F1525/Active site: Cys (phosphocysteine intermediate) #status predicted  
F1531/Binding site: substrate phosphate (Arg) #status predicted

## Query Match

9.6%; Score 1150; DB 2; Length 1615;

Best Local Similarity 24.9%; Pred. No. 6.4e-49;  
Matches 438; Conservative 266; Mismatches 611; Indels 442; Gaps 74;

QY 683 RLKRGPP---EKPNGIIIAEYVLYKNIDTLYMKNTSTTDIILRNLRPHTLYNISVRSYTR 739  
Db 90 RIDYSPFPFPFNITIPASDI-----GKDKFSRALPGTEYNFWL--YYTN 134  
QY 740 FGHNQVSSLLSVRSTETVPSAPENITYKNISGGEIELSFLPPSPN-----GIKKY 793  
Db 135 STHQELTWVNIITTA---PD-PPANLSVQLRSKSAFTWRPPGSGRYSGFIRVLGLT 190  
QY 794 TTYLRES---NGNEERTINTTSITQNIKV-LKKYQYIIIEVSASTLKGSGVRSAPISILT 849  
Db 191 DLPFRSISLEGNETLQLSAKELTPGGSYQOAYISVY-----QKGSVATSRN 239  
QY 850 BEDAPDSPQDFSVKQLSGVTVKLSWQPPLEPENGIIYYTYVYVWNRSSLKTNVTETSL 909

Db 240 FTKPNTPK-FIVWRNETTLVLMQPP-PAGIYTHRVSTPDDAIQSVLYVER--- 294  
QY 910 LSDLDYVVEYSAYVASTAATRGDKTGSNIISFOTPEGAPDPKPVYYANLSSSSILFW 969  
Db 295 -----EGEPPGPAQAAP----- 306  
QY 970 TPKSKNGII--QYVSVYVNTSGTFMQNFTLHETLNDPDMNTVSTIIDKLTIFSYTF 1026  
Db 307 -----KGLVPGRHY-----NISQTVSEB-ETSSVFTTARYLT----- 338  
QY 1027 WLTAHSVGNKSSDIIIEVYTDQIDPEGFVGNLTYES--ISSTAINVSVVPAQNGLV 1084  
Db 339 -----VPER-VLNVTFDEAYTSSFRVREPPRTYSBED 372  
QY 1085 FYVVSIILOQTPRHVRPPLVYERSIYFDNLEKYTDYILKITPSTKGFSDTYTAQLYIK 1144  
Db 373 AYQV--MLSTSRIIFNVPAANGDSVFD-----YSIDL-----BPG-----RTYEVVVK 415  
QY 1145 TEEDVPTSP-----INTFKMLSTSVLLSWDPV--KPGALISYDLTLQ 1189  
Db 416 TIADNVNVPASGEVLRPRPVASLGFLDSDRNALHISWEPAETGRQDSYRISYEQTN 475  
QY 1190 GPNENYSFITSNDVNIILELSPFTLYSEFAAAR-----TRKGLGP-SSILFFYDES 1241  
Db 476 ASEVPAPFVAESQITTNLTETLDSLLAGRYLIAVQALSGVASNASDITRYTRPAA 535  
QY 1242 PLAPPQNLTLNCTDFVWLKWSPLPGGIVKVS-----FKIHEHETDIYKNSGPK 1297  
Db 536 PLI--QELASID--QGLMLSWSD-----VNSRQDRYEVHYQNGTREERTWATNE 582  
QY 1298 TEAKLVGLSPVSYSTRVSAFTKVGNGQPSNVVKTQESVDVQVQNMOCWATSQSVL 1357  
Db 583 TSLTIHLPFGSGYEVKVAH--ISHGVRSSEPHSYF--QAVFPKPPQNLTLQVHTNLV 637  
QY 1358 VKWDPPKANGIITQVMTVVERNSTKVP-----QDHMYTFIKLLANTSYV 1403  
Db 638 LHMQAPEGSD--FSEYV--RYRTDASPQRISGLHENEARKDMHY-----GERYL 685  
QY 1404 FKVRASTAGDESTCHVSTLPETVPSVPTNIAPSDVOSTATLWIRPDTILGFQNY 1463  
Db 686 VQV--NTVSFGVESP-HPELVNTWPPQVSVNVPLVDSRNLTLEWPRD---GHVDYF 738  
QY 1464 KITTLQRAQCKEW--ESECEVEYKIQVLYEAHLTEETV---YGLKKFRWYRFOVAAS 1517  
Db 739 TL-----KWPTEDEEDRVEFKVNTQLED--LSSPSVRIPEDLSPGQRYRFEVQAS 787  
QY 1518 TRAGYGNASNWISTKLP-----GPPDGPENVHVAT-SP-----FS- 1554  
Db 788 SN-GIRSGTHTLSTRMTPLIQSDVFIANAGHEQGDETITLSTYPTPADSTRFDIYRF 846  
QY 1555 -----ISISWSEPAVITGPTCYLI-----DV 1575  
Db 847 GPTIKDKEKLANDTERKLSFGLTGKLVNVTWVSGVASLPVQVRVRLHPLPISDL 906  
QY 1576 KSV-----DNDEFNISFKSNEE-----NKTBIKOLEIFTRY 1609  
Db 907 KAIQVAAREITLHTAPAGEYTDPELQYLSADEEAPQLLQVNTKVTITLQGLRPVHNT 966  
QY 1610 VITATGNISAAVVGKSAEMIVTL--ESAP-----KDPNNMTFOKIPDEVTK 1659  
Db 967 FTWVVRSGSI-----QGTDFADVSVSTLMRSSAPISASYQTLTAPPKGVDFQ-PSDVQ 1020  
QY 1660 FQLTF--LPPSQPNQNIQVQALVYREDPTAVQIHNLSIIQKTNFTVIAMLEGLKGGH 1716  
Db 1021 GEVTFEWSLEPAEQHGDIFRITCONADDAADVSEFPV-----NATQKIDGLVPGN 1075  
QY 1717 TYNISVAVNSAGAPKPMRIIMDIKAPARKTKPTPYDATGKLLVSTTITIRMPIC 1776  
Db 1076 HYIFRQAKSALGYGAERHIQTMPI LAPVPEPSVTPL-----EVSRTSSIIETISFROG 1130  
QY 1777 YSDDDHGPKNQV--VLATETGAQHDG--NYTKWD--AY-----FNKARYFTNEGPNPP 1827

Db 1131 YFSNAHGMVRSYTIITAEADVGIASGLEMPMSQDVQAYTVWLPHYQAIEPY-----NPF 1183  
QY 1828 CT-EGTKPFGNEEIIYGADNACMTIPGNEDK-----ICNGPLKPKKQYLYFKFRATIMQG 1882  
Db 1184 LTSNGSRKSLAEHLTIGTANC-----DKHQAGYCNGLPRAGTYYRIKIRAFATDEDK 1236  
QY 1883 FTDSDYSDVKTUGELSERTEIILSVTLCTILSLITLGTAFAPARIQKQEGGTYSP 1942  
Db 1237 FDTVTYSSPITT-----ERSDTVIVAAT---VSAVLLVAMVLV-----VYCQ 1276  
QY 1943 QDAEIIDTTLKL-----DQLITVADLEKDBERLTRPISKSKSFLQHVBEELCTNNNLKFOEER 1998  
Db 1277 HRCOLIRRAKLARMQDELAALPEGYITP---NRPHVKDFSEHYIRMSADSFSEEP 1333  
QY 1999 SELPKFLQDLSSTDADLPWNRKRPNNKPNKNNRVKLIADASVPGSDIYINASYG 2058  
Db 1334 EELKHVGRDQACSFANLPCNRPKNRFTNLPY-DHSRFLQLQPVDDDDGSDIYINAMPGH 1392  
QY 2059 LCPNEFIATQGPLPGTGVDFWBMVWETRAKTLVMTQCEKGRIRCHOYWPEDNKPVTVP 2118  
Db 1393 NSPREFIVTQGPFSHSTREEFWRMCWESNRAIVMLTRCFEKGREKCDQYWPVD-RVAMFY 1451  
QY 2119 GDIVITKLMEDVQIDMTIRDLKIERHGDQWTVRQCNFTAPWEPHGVGPENSAPLHFVKLVR 2178  
Db 1452 GDIKVQLIIDTHYHDSISEFWVRNCSERIMRHFHTTWDPGVEPPPLSLVRFVRAFR 1511  
QY 2179 ASRAHDTTTPMIVHCSAGVGTGVFIADHLTQHINDHDFVDIYGLVAELSERCMVQNL 2238  
Db 1512 DVIQDTRMFIIVHCSAGVGRSGTFFIALDRLQIHHSKDYVDIFGIVFAMRKEVFMVQTE 1571  
QY 2239 AQVIFLHQICILLNSK 2255  
Db 1572 QQYVCIHQCLLAVLECK 1588  
  
RESULT 9  
A49502  
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type 4E, splice form A precursor  
C:Species: Drosophila melanogaster  
C>Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
C:Accession: A49502  
R:Oon, S.H.; Hong, A.; Yang, X.; Chia, W.  
J. Biol. Chem. 268, 23964-23971, 1993  
A:Title: Alternative splicing in a novel tyrosine phosphatase gene (DPTP4E) of Drosophila  
A:Reference number: A49502; MUID:94043220; PMID:8226938  
A:Accession: A49502  
A:Molecule type: mRNA  
A:Residues: 1-1767 <OON>  
A:Cross-references: UNIPROT:Q9M4F5; UNIPROT:Q24495; GB:L20894  
A>Note: authors translated the codon ATA for residue 1715 as Leu  
C:Genetics:  
A:Gene: ptp4E  
A:Cross-references: FlyBase:FBgn0004368  
C:Superfamily: protein-tyrosine-phosphatase, receptor type 4E; fibronectin type III rep  
C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; recep  
F:1254-1270/Domain: transmembrane #status predicted <TM>  
F:1271-1767/Domain: intracellular #status predicted <INT>  
F:1353-1573/Domain: protein-tyrosine-phosphatase homology <PTP>  
F:1525/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:1531/Binding site: substrate phosphate (Arg) #status predicted  
  
Query Match 9.6%; Score 1150; DB 2; Length 1767;  
Best Local Similarity 24.9%; Pred. No. 7.4e-49;  
Matches 438; Conservative 266; Mismatches 611; Indels 442; Gaps 74;  
  
QY 683 RLKWSPP---EKPNGLIIAYEVLVYKNIDFLYMKNTSTTDIIILNRLPHLYNISVRSYTR 739  
Db 90 RIDYSPFPFGPEPNTTIPASDI-----GKDIKFSRALPGTEYNFWL--YYTN 134  
QY 740 FGHGQVSVLLSVRSETVPSAPENITYKNISGGEIHSFLPPSPN-----GIKKY 793  
Db 135 STHRQLTWTVNITTA---PD-PPANLSVQLRSSKSAFITWRPPGSGRYSGFRIVLGLT 190



F:1554/Binding site: substrate phosphate (Arg) #status predicted  
F:1839/Active site: Cys (phosphocysteine intermedate) #status predicted  
F:1845/Binding site: substrate phosphate (Arg) #status predicted

Query Match 9.5%; Score 1140; DB 2; Length 1907;  
Best Local Similarity 25.5%; Pred. No. 2.6e-48;  
Matches 424; Conservative 267; Mismatches 606; Indels 364; Gaps 66;

QY 750 LSVRTSETPDAPENITYKNISGIEILSFLPPSPNGIHKYTYLYLKRSGNEERT-- 807  
DB 223 LYVRVRVAPRFSILPMSHEIMPNGVNTITCVAGSP-----MPYVWMOGAEDLTPE 275  
QY 808 -----INTTSITONIKVLKXYQYIEVSASTLKGEGVRSAPISILTEADAPDSPQDP 861  
DB 276 DOMEVGRNLELT-DVKDSANYTC-----VAMSSL-GVIEAVAQITVKSIFKAPGTP- 325  
QY 862 SVKQLSGVTVKLSQWPPLENGIILYVTVVWNR-----SLKTINVTSTLSLESLDYN 916  
DB 326 VVTENTATSTVTWDSG-NPDP-VSYVYIEYKSKSQDPYQIKE-DITTRYSIGLSFN 382  
QY 917 VEYSAYVTASTRFGDKTGSNIISFOTPEGAPDPKPPQVYVYANLSSSIIILFWTPPSKPN 976  
DB 383 SEYEIWSVANSIGQPPSPESVVT-RTGEQAPASAPRNQARMLSATTMIWQEBPVEN 441  
QY 977 GIIQYYSVYRNTSGTFMQNFTLHETNDPDMTSTIIDKLTFISYTYFWLTASTSVGN 1036  
DB 442 GLIRGRVYVTEPEHPVGNWQKHNV-DPSLLTFTVGSLLDEET-----YTVRVLAFTSVGD 496  
QY 1037 GNKSSDIIEVYTDQDIPGEGVGNLTYESISTAINSVWPPAPNGLVFVYVYLLIQPT 1096  
DB 497 G-PLSDPIQVKTQGVG-PQPMRLBAKSETSIGLSWAPROESVIKY---ELLFREGD 551  
QY 1097 RHRVPLVYERSIYE--DNLEKYTDIILKITSTEGFSDTYTAQYIKTEBEDVPETSP 1154  
DB 552 RG-REVGRTFDPTTAFVWEDLKNTFYAFELAARSPOGLG-AFTAVVHQRTLOAKSAPP 609  
QY 1155 IINTFNKLSSTSVLSW--DPPVKPNGAISYDLTLQ-----GNENYSFITSNYYI 1205  
DB 610 QDVKCTSLASTALVSWRPPPTHTNGALVGVSVYRPLGSEDDPKVNNIPPTTQIL 669  
QY 1206 LEELSPTLYSFAAARTKGLP-SSILFFYTVDESVPAPQNLTLINCTSDFWLWKS 1264  
DB 670 LEALEKWTYRVTAVAYTEVGPSPSSVVRVDEDPSPAPRKVEAEALNATKRVLR 729  
QY 1265 PSLPLG---GIVKVSFKIHEHETDIYKNISGPKT-----EAKLVGLE 1306  
DB 730 -SFTPRQHQIRGYQ-----VHYVRMEGAERGAPRIKDIMLADAQEWITNLQ 778  
QY 1307 PVSTYSIRVSAFTKGVNGNQFSNVKFTQESVPDVV-----QNMOCMATSWQSVL 1357  
DB 779 PETAYSITVAAYTMKDG-----ARSKPKVWTKGAVLGRPTLSVQQTPEGSL 827  
QY 1358 VKWDDP-KKANGIITQYMTVERNSTKVPQDHWYTFIKLLA-----NTSYVFK 1405  
DB 828 ARWEPGDAEDPVLGYRLQFGR-----EDRAPATLEAAWERRFAAPAHKGAITYVR 880  
QY 1406 VRASTAGSGDESTCHVSTLPETVPSVPTNI--AFSDVOSTSATLWIRPDIT--LGYFQ 1461  
DB 881 LAARGGGLGEEAALIS-IPEDAPGFPQILGAGNVAGSVLLKWLPLPVAERNGAIL 939  
QY 1462 NYKITTLQRAQCKEWESECEVYQIYLYEAHLTEETVYGLKFRWYRFVAASTNAG 1521  
DB 940 KYTVSVREAGAPGATETELAAQAQ-----PGAETALTGLRPLPETGYELVRAHTRRG 993  
QY 1522 YGNASNIWSTKLPGPPDP-----PENHVATSPFSISISWSEPAVITGPTC 1570  
DB 994 -----PGPSPPLRYRLARDPVSPKPNFKVMKMTSVLLSWFEFDVNSPTP 1040  
QY 1571 YLIDVKSVDNDEBNFISFKSNEENKTIE--INKDEIFTRYVVVITAFCTNISAAY--VSG 1626  
DB 1041 YKIQYNGLTLTD-----VDGRITTKLLIHLKPHFTYFNVFLTRNGSLGGLQQTVA 1090  
QY 1627 KSSAEMIVTLESAPKDPNNMTFKIPDEVTKFQLTFLPPSQPNNGNIQVQALVYREDD 1686

DB 1091 RTAFNMLSKSPVAPK--PDN-----DGFIVVYLPDQG-----S 1122  
QY 1687 PTAVQIHNLSIIQKTNFTFIAM--LEGLKGHTYINISVAVNSAGAGPKVPMRITMDIKAP 1745  
DB 1123 PVTVQ-----NYFIVMVLAKSRG----- 1142  
QY 1746 ARPQKPTPIYDATKLLVTSTIIRMPICVYSDHGGIKNVQVLATATGAQHDGNVTK 1805  
DB 1143 -----QFVLLGSPEDMDLEBLIQDISLQORH-----V 1171  
QY 1806 WYDAFVKARPFTNEGPNPCTEGTKFS-----GNEEYITIGADNACMIPGNE 1857  
DB 1172 RHRQLEVRPXIA-----ARFSLPVAVFHNGKQY--GGFNRGLEPQH-- 1215  
QY 1858 KICNGPLPKQYLKFKFRATNMGQ--FTDSYSDPV-----KTLGSEGLSERTVEIIL 1908  
DB 1216 -----RYVLVFLAVLQKNEPTFAASPFDPQLDNPDPQPIVDG--BEGLIWI 1262  
QY 1909 SVTLCLISILIGTAIFAPARIQKQEGGTVSPQDAEIIDTKLKLDQILITVADL---EL 1965  
DB 1263 GPVLAUVFIICIVIAILLY-----KNK-----PDSKPKDSEPRTKCLLNADLAPHHP 1310  
QY 1966 KDERLTR-----PISKKSFLQHVLEELCTNNNLKQEFSESL---PKELQDLS 2009  
DB 1311 KDPVEMRRINFOTPGMLSHPPITPITDMAHMERLXANDSLKSQBYESIDPQOQQFWHS 1370  
QY 2010 STDADLPWNRACKRPNFKPNNNRVKLIADASVPGSDYINNASYISGYLCPNEFIATQG 2069  
DB 1371 NLEA-----NKPKNRYANVIAY--DHSRVILQPLEGMSDYINANYVDGVRQNAVYATQG 1425  
QY 2070 PLPGTVGDFMRWVETRAKTLVMLTQCFEKGIRCHQWPNEDKNPVTVFGDIVITKMBD 2129  
DB 1426 PLPETGDFMRWVQORSATVVMVTRLEKSRKCDQWY--NRGTETTYGFIQVT--LLDT 1482  
QY 2130 VOI-DWTIRDLKIERHDC--MTVQCQNFETANPEHGVSPENSAPLHFKVLVRSRAHDTT 2186  
DB 1483 MELATFCVTRTSLHNGSGKREVRHFQTAWPDHGVPEYPTFLAFLRRVKTCPNPDAG 1542  
QY 2187 PMVHCSAGVGTGVFIADLHLOTHNDHDFDIYGLVLAELSERMCMQNLQAYIFLHQ 2246  
DB 1543 PIVHCSAGVGTGTCFVIDAMLERIKTEKTVDVTVGHVTLMSQRNMVQTEDQYGFIIHE 1602  
QY 2247 CILDLLSKNGSNQPCFVNYSAQKMSLDLAMEGDEVELEWE 2287  
DB 1603 ALLEAVSCGTEVPARSL-YTVIQKLAQVEPGEHVTGMELE 1642

## RESULT 11

D41214

protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type 10D, short splice form precursor  
C:Species: Drosophila melanogaster  
C:Date: 28-May-1992 #sequence\_revision 28-May-1992 #text\_change 09-Jul-2004  
C:Accession: D41214; A41215  
R:Yang, X.; Seow, K.T.; Bahri, S.M.; Oon, S.H.; Chia, W.  
Cell 67, 661-673, 1991  
A:Title: Two Drosophila receptor-like tyrosine phosphatase genes are expressed in a sub  
A:Reference number: A41214; MUID:92034988; PMID:1657401  
A:Accession: D41214  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-1557 <YAN>  
A:Cross-references: UNIPROT:P35992; GB:M80465  
R:Tian, S.S.; Tsoulfas, P.; Zinn, K.  
Cell 67, 673-685, 1991  
A:Title: Three receptor-linked protein-tyrosine phosphatases are selectively expressed  
A:Reference number: A41215; MUID:92034989; PMID:1657402  
A:Accession: A41215  
A:Molecule type: mRNA  
A:Residues: 1-904, 'L', 906-1125, 'Q', 1126-1165, 'YR', 1168-1171, 'A', 1173-1215, 'L', 1217-1456  
A:Cross-references: GB:M80536; MUID:9158644; PIDN:AAA28952.1; PID:9158645  
C:Genetics:  
A:Gene: FlyBase:Pt910D







FEBS Lett. 282, 285-288, 1991

A:Title: Expression of a truncated protein-tyrosine phosphatase mRNA in human lung.

A:Reference number: S15818; MUID:91243813; PMID:1645282

A:Accession: S15818

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1872-1911, 'VHMLQK' <VR1>

A:Accession: S15819

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1872-1997 <VR2>

C:Genetics:

A:Gene: GDB:PTPRB; PTPB

A:Cross-references: GDB:127352; OMIM:176882

A:Map position: 12q15-12q21

C:Superfamily: protein-tyrosine-phosphatase, receptor type beta; fibronectin type III re

C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; recept

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-1997/Product: protein-tyrosine-phosphatase, receptor type beta #status predicted <M

F:23-1625/Domain: extracellular #status predicted <EXT>

F:1626-1642/Domain: transmembrane #status predicted <TM>

F:1643-1997/Domain: intracellular #status predicted <INT>

F:1727-1952/Domain: protein-tyrosine-phosphatase homology <PTP>

F:1904/Active site: Cys (phosphocysteine intermediate) #status predicted

F:1910/Binding site: substrate phosphate (Arg) #status predicted

Query Match 9.2%; Score 1107; DB 1; Length 1997;

Best Local Similarity 23.8%; Pred. No. 1.2e-46;

Matches 519; Conservative 324; Mismatches 790; Indels 550; Gaps 101;

373 YIAAETSAGT-----GPKSNISVFTPPDPCA-----VFIDLAEVEST 411

29 FTLAESKASHSVSIQWRILGSPCNFLSYSSDTLGAALCPTFRIDNTTYGCLQDLQA- 87

412 QVRITWKQPNQNGIINQYRVKLVPEPTGIIENLLTLLGNBYINDPMAPRINIVPEMV 471

88 -----GTIYNFKIISLBERTVWLQ-----TDLPAPARFGVSK- 120

472 GLYEGSAEMSSDLHSLATIYNHSPDKNFPARNRAEDQTSVVVTFNNOVITDIAEQLS- 530

121 -----EKTSTGLHVMWT-----PSSGKVTSEVQLFDENNKIQGVQIQUESTS 164

531 ---YVIRRLVPFTEHMSVSAFTIMGEGPTVLVSRTR-QQVPSIIKIINYNKISSSIL 586

165 WNEYTFNLTAGSKYNIATVVS-----GKRFSFVYVNGSTVPSPKVDIGI-STRANSL 219

587 LYWDPPPEYNGKITHYIYAMELDTNRAFQITID---NSFLITGLKKYTKYNMRAAST 643

220 ISW---SHSGGNVRYRL--MLMDKGLVHGGVVDKHAQTSYAFHGLSPGYLYNLTVMTEA 274

644 HGGESSLSENDIFVTSDEPSSQDVEIDVTAD-----EIRLKWSPPEKNGIIAY 699

275 ---AGLQYRWKLVRT-----APMEVSNLKVTDNGSLTSLKVKW---QRPPGNVDYSY 320

700 EVLYKNIDITLYMKNTS-----TTDILRLNRLPHTLYNLSVRSYTRFGHGNQVSSLL 750

321 NI-----TLSHKGTIKESRVLAPWITETHFRELVEGRLYQVTV-----SCVSGEL 365

751 SVR---TSETVPDSPAENITYKNISSGEIELSPSPNGIHKYIYLKESNGNEERT 807

366 SAQKMAVGRTFPDKVANLEANNRMRLSVLVSPPA---GDWEQYRILL----- 412

808 INTTSITONIKVLKKYTOYIIE-----VSASTLKG-----EGVRSAPISIL 848

413 FNDSSVLLNITVKEETQYVMDPTGLVPGRQYVEVEIVESGNLKNKNSERCQG-RTVPLAVL 471

849 TEEDAPDPSQDFSVKQLSGVTVKLSQWQPL---EPNGIILYVTVVNRSSL---KTIN 902

472 -----QURVXKANTSLUSIMWQTPVAWEK-----YIISLARDLLLLHKSLUS 514

903 VTETSLESLDLYNVEYSAYVTASTRFGDKTGKTSNIISFTQPEGAPSDPPKDVYANLS- 961

515 KOAKEFTFTDLVGRKYMATVTSIS--GDLKNSSSVKGRTPVPAQV-----TDLHVAQGM 567

962 SSSIIIFWTPPSKPNGIIOYISVYVYRNTSGTFMQNFTLHETLNDPDMTV--STIIDKLT 1019

568 TSSLFTNWT---QAQGVFEYQV-----LLIHE-----NVVKNESISSETS 606

1020 IFSYY-----TFWLTAHSVGNKNSDIIEVYTDQDIPGEGFVGNLTLYESISSTAINVSW 1074

607 RYSPHLSKSGSLYSVVVTVTSGGSSRQV--VVEGRTVPSSVSGVTVNSGRNDVLSVSW 664

1075 VPPAQPNGLVFTYYSL-----ILQOTPRHVPPLVTVYERSIYEDNLEKYTDYILKITST 1129

665 L--VAPGDVDNYEVTLSHDKVQSL-----VIAKSVRECSFSLTPGRLYTVTIT--T 714

1130 EKGFSDTYTAOLYIKTERDVPETSPINTFNKLSTSVLLSWDPDPKNGGAIISYDLTLQ 1189

715 RSGKVENHS-----FSQERTVPDKQGVSVNSARSDDLRSW--VHATGDFDHVEVIK 767

1190 GPN---ENYSFITSNYYIIEELSFPTLYSPFAAART-----RKLGPSSILFFYTDDES 1240

768 KNNFIQTKSPKSENECVFQVLVGRLYSVTVTTKSGQYEANEQNG-----RT 817

1241 VPLAPPQMLTLINCTSDFWLKWSPSLPGGIVKVKYSFKIHEHEHETDIYKKNISGFKTEA 1300

818 IP-EPVKDLTLRNKSTEDLHVTWGA---NGVDQYEQIQLLFNDMKVFPFPHLVNTATYEY 873

1301 KLVLGPEVSTYSIRVSAFTKVGNGQFNSVVKFTTQESVDPVQVNMOCMAT--SWOSVLVK 1359

874 RPTSLTPGRQKILV--ITISGDVQOQSAFIEGFT---VPSAVKNIHISPGATDSLIVN 927

1360 WDPPKKAAGIITQYVMVTVERNSTKVSQ---DHMY--TFIKLLANTSVVKVRASTAGE 1414

928 WTP---GGGDVDSYTVSAFRHSQKVDSTIPKHPVEHTFHRLEAGEQYQIMI-ASVSGSL 983

1415 GDESTCHVSTLPETVPSVPTNIAFSDVOSTSATLTWIRPDTILGYFQYVKI----- 1465

984 KNQINVVGRTVPASVQGVADNAYS---SYSLIVSW---OKAAGVAERYDILLTENGIL 1037

1466 ---TTLRAQCKEWESECEVEYQKIQYL-----YRQ-----VAAST 1518

1038 LRNTSEPAATTKQHKEDLTGKKYKIQILTVSGGLFSKEAQTEGRTVPAAVTDLRITENS 1097

1501 VYGLKKFRW-----YRQ-----YRQ-----VAAST 1518

1098 TRHL--SPRWASBEGELSWNIFLYNPQNLQERAQVDPLVQSFQNLQGRMYKQWIVT 1156

1519 NAGYGNASNWISTKTLPGPPD---GPPENVHVATSPESISISWSEPAVITGPTCYLLDV 1575

1157 HSGELSNESFPGRTVPASVSHLRGSRN-----TTDSLWFFNWS-PA----- 1197

1576 KSDVNDDEFNISPIKSEENK-TIEIKDLEIF-----TRYSVVITAFITGNISAAVVEG 1626

1198 -SGDFDFTELLILYNPGTKKENWKDKDTEWRPQGLVGRKKVLMVWVTHSGDLS-----N 1251

1627 KSSAEMIVTTLESAPKPPNNMTFOKI-----PD--EVTKFQLTFLPPS----- 1668

1252 KVTAES-----RTAP--SPPLMSFADIANTSLAITWKGPDPDWDYDNDPELQWLRDALTV 1305

1669 -OPNGNIQVQALVY-----RE 1684

1306 FNPYNNRKSEGRIVYGLRPGRSYQFNKTVSGDSWKTSKPIFGSVRTKPKDKIQLNLCRP 1365

1685 DDPTAV-----QIHNLSIIQKWTFTVIAMLEGLKGGHTYNI 1720

1366 QNSTAIACSWIPDSDPDGYSIECRKMDTQEVFSPKLEKESLLNIMM--LVPHKRYLV 1423

1721 SYIAYNSAGAKVPMRITMDIKAPARKTKPTPIYDATGKLLVLTSTTITTRMPICYSD 1780

1424 SI-KVQSGAGTSEVVEDST--ITMDRPPPPPPHVRNEKDVLSKSSINFVNCWSFSD 1480

1781 DHGPIKNQVOLTATGA-----OHGDNVTKWTDAYENKA--RPYFTNEGFPPPPCTE 1830

1481 TNGAVKYTVVVVREADGSDDELKPEQQHP---LPSYLEYRHNASIRVYQTN--YFASKCAE 1535

1831 GKTKFSGNEBIYIIGADNACWIPGNE- ----KICNGELPKKYOYLPKFA-TNI-----WG 1881  
Db 1536 NPNNSNSKFNK-IGAME-SLGKRDFTQOKFCGPKLPHHTYAIRISIRAFQLEDLK 1593  
Qy 1882 QFTDSYSD-----PVKTLGEGLSERTVEIILSVTLGILSIIILGTAIFAFARIQOKKE 1936  
Db 1594 EFTKPLYSDFPFLPITTESEPLF-GAIEGV-SAGLFLIGMLV---AVVALLICRQKVSH 1648  
Qy 1937 GGTYSPODABIIDTKLKLQILIV-ADLEKDLRLTR-PLSKSFLOHVEELCTNNLKF 1994  
Db 1649 G-----RERPSARLSIRDRPLSVHLNGLQGNKRTSCPIKINQFEGHFKLQADSNYLL 1703  
Qy 1995 QEFSELPKFLQDLSSTADLPWNAKRPNNKPNVNNRVKLIADASVPGSDYINASY 2054  
Db 1704 SKYTEELKQVNRQSCDIALPENRGNKRYNNILPY-DATRVKLSNVDDPDCSDYINASY 1762  
Qy 2055 ISGYLCPNEFIATQGPLGTGVGFWRMVWETRAKTLVMTLQCEKGRIRCHQWPDNKP 2114  
Db 1763 IPGNPFREYIVTQGPLGTGDFWKMVWQNVHNVMTQCEKGRVCKDHWPADQDS 1822  
Qy 2115 VTVFGDIVITKLMEDVOIDWIRDLKI---ERHGDCTVTRQCNFTAWPEHGVSPENSAPLI 2171  
Db 1823 L-YGDLILQWLSVLPWTIREFKICGEEQLDAHRLIRHFHYTVPDHPGVETTSQSLI 1881  
Qy 2172 HFVKLYR--ASRAHDTTPMIVHCSAGVGRVGVFIALDHLTOHINDHDFVDIYGLVABLRS 2229  
Db 1882 QFVRTVRYDYNRSPGAGTVPVHCSAGVGRGTGTFTIALDRILQDLSKDSVDIYGAVHDLRL 1941  
Qy 2230 ERMCMQNLQAQYIFLHCILDL 2252  
Db 1942 HRVHMVQTECYVYLHCQVRDVL 1964

RESULT 14  
S46217  
protein-tyrosine-phosphatase (EC 3.1.3.48) type sigma precursor - rat  
N;Alternate names: leukocyte common antigen-related phosphatase  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 07-May-1995 #sequence revision 03-Nov-1995 #text change 09-Jul-2004  
C;Accession: S46217; S51174; A49104  
R;Zhang, W.R.; Hashimoto, N.; Ahmad, F.; Ding, W.; Goldstein, B.J.  
Biochem. J. 302, 39-47, 1994  
A;Title: Molecular cloning and expression of a unique receptor-like protein-tyrosine-phosphatase from rat leukocytes  
A;Reference number: S46217; MUID:94347119; PMID:8068021  
A;Accession: S46217  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-1863 <ZHA>  
A;Cross-references: UNIPROT:Q64605; EMBL:L11587  
R;Goldstein, B.J.  
submitted to the EMBL Data Library, February 1993  
A;Reference number: S51174  
A;Accession: S51174  
A;Molecule type: mRNA  
A;Residues: 1-1788, 'G', 1790-1863 <GOI>  
A;Cross-references: EMBL:L11587; NID:9205134; PID:AAC37656.1; PID:9205135  
R;Van, H.; Grossman, A.; Wang, H.; D'Eustachio, P.; Mossie, K.; Musacchio, J.M.; Silverman, J. Biol. Chem. 268, 24880-24886, 1993  
A;Title: A novel receptor tyrosine phosphatase-sigma that is highly expressed in the nervous system  
A;Reference number: A49104; MUID:94043351; PMID:8227050  
A;Accession: A49104  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: nucleic acid  
A;Residues: 1-596, 'R', 598-603, 'I', 967-1788, 'G', 1790-1863 <VAN>  
A;Experimental source: brain  
A;Note: sequence extracted from NCBI backbone (NCBIP:139669)  
C;Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology; cyg  
C;Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester hydrolase  
F;1-26/Domain: signal sequence #status predicted <SIG>  
F;27-1863/Product: protein-tyrosine-phosphatase #status predicted <MAT>  
F;149-209/Domain: immunoglobulin homology <IMM1>  
F;246-300/Domain: immunoglobulin homology <IMM2>

F;318-400/Domain: fibronectin type III repeat homology <FN3>  
F;413-499/Domain: fibronectin type III repeat homology <FN3B>  
F;511-592/Domain: fibronectin type III repeat homology <FN3C>  
F;1244-1863/Domain: leukocyte common antigen cytosolic domain  
F;1331-1552/Domain: protein-tyrosine-phosphatase homology <PTPI>  
F;1504/Active site: Cys (phosphotyrosine intermediate) #status predicted  
F;1510/Binding site: substrate phosphate (Arg) #status predicted  
F;1795/Active site: Cys (phosphotyrosine intermediate) #status predicted  
F;1801/Binding site: substrate phosphate (Arg) #status predicted

Query Match 9.0%; Score 1079; DB 2; Length 1863;  
Best Local Similarity 25.2%; Pred. No. 2, 7e-45;  
Matches 416; Conservative 262; Mismatches 582; Indels 394; Gaps 66;

Qy 750 LSVRTSETVDSAPENITYKNISSGEIELSFPPSPNGIIKYTYTYLKRSGNEERT-- 807  
Db 223 LVVRVRVAPRPSILPMSHEIMPNGVNITCVAVSP-----MPYVKKMQGAEDLTPE 275  
Qy 808 -----INTSLTONIKVLKYYTYIIEVSASTLKGEGVRSAPISILTDEADPDSPPQDF 861  
Db 276 DMPVGRNVLELT-DYKDSANYTC-----VAMSSL-GVIEAVAQITVKSLPKAPGTP-- 325  
Qy 862 SVKOLSGVTVKLSWQPPLEPNGIILVYTVVWNRS-----SLKTINVTTLSLESLDYN 916  
Db 326 VTENTATSIITWDSG-NPDP-VSYVYIEYKSKSDGPGYQIKE-DITTRISIGGLSPN 382  
Qy 917 VEYSAYVTASTRFGDKTGSNIISFQTEGAPSDPKPKVYVYANLSSSSIIILFWTPPSKPN 976  
Db 383 SEVEIWSAVNSIGQPPSPESVVT-RTGQAPASAPARNVQARMLSATTMIVQWEEFVEPN 441  
Qy 977 GLIQYVYVYRNTSGTFMQNFTLHETNDFDMNTVSTIIDKLTFISYTYFWLTASTSVGN 1036  
Db 442 GLIRGYRVYTYMEPEHPVGNWQKHNVDLSLLTTVGSLEDET-----YTVRVLAFTSVGD 496  
Qy 1037 GNKSSDIEVYTDQDIPGFGVGNLTVESISSAINVSWVPVPAQNGLVVYVYVLSLIQOOP 1096  
Db 497 G-PLSDPIQVKTOQGV-PQPNVLRARAKSETSIGLSWAPQESVIKY---ELLFREGD 551  
Qy 1097 RHRVPLVYERSIYF--DNLEKYTDYILKITPTEKGFSDTYTQALYIKTBEDVETSP 1154  
Db 552 RG-REVGRFTDPTAFVVEDLKPNTHEYAPLAARSQGLG-AFTAVVCQRTLOAKPSAPP 609  
Qy 1155 INTFNKLSSTVLSWDPVVKPNGAISDYDTLQGNENYFITSNDVNIIEELSPPFL 1214  
Db 610 QDVKCTSLRSTAIL-----ILLEALEKWT 634  
Qy 1215 YSFEAAARTRKGLP-SSILPFTYDPSVPLAPPQNLTLINCTSDFWLWAKWSPPLPG--- 1270  
Db 635 YRTAVAYTEVGGPESPVPVVRDEDPSPAPRKRVEABALNATAIRVLWR-SPTPGROH 693  
Qy 1271 GIVKVSFKIHSHTDTIYKNISGFKTEAK-----LVGLEPVSYSI 1313  
Db 694 GOIRGYQ-----VHYVMEG--TEARGPPRIKDIMLADAQEWITNLQETAYSI 741  
Qy 1314 RVSAFTKVGNGQFNVVKTQESVDDV-----QNMCMATSWQSVLVKWDPP- 1363  
Db 742 TVAAVYTMKGDG-----ARSKPKVVTGKAVLGRPTLSVQQTPEGSLIARMEPPA 790  
Qy 1364 KKANGIITQYMYTVERNSTKVSQPDHMYTFIKLLA-----NTSVYFKVRASTSA 1412  
Db 791 DAEDPVLGYRLQFGR-----EADAPATLELAWERRFAAPAHKATYVRLAARGA 843  
Qy 1413 GEGDESTCHVSTLPETVPSVPTNI--AFSDVQSTSATLTWIRPDTILGYFQNYKITTLQR 1470  
Db 844 GLGEASAALS-IPEDAPRGFPQIILGPNAGNSAGSVILRWLPVPAEGNGAIKYTVSVR 902  
Qy 1471 AOKCKEWSBECVYQKTYLYEHLTBETVYGLKFKFRWYRQVAASTNAGVGNASNWS 1530  
Db 903 ----BAGTPGPATETELAAAAQGAETALTQGLRPETAYELRVRAHTRG----- 949  
Qy 1531 TKTLQCPDGP-----PENVHVAVSPFSISISWSSEPAVITGTCYLVLDKVSVD 1579  
Db 950 ----FGPFPPLRYRLARDPVSPKPKVIMKTSVLLSWEPFDNYSPTPKYIQNGLT 1005

**Cy** 1580 NDEFNIFSIKSNENKTIE--IKDLEIFTYYSVVITAFIGNISAAY--VEGKSASAMIVT 1635  
**Dg** : : : | : | : | : | : | : : : :  
1006 LD-----VDGRTKKLILHLKPHTFYNFLVLTNRSGSLGLQOQTARTAFNMLSG 1055  
1636 TLESAPDKPNNMTFKQIPDEVTKFQLTFPPSPNGNIQVVALVVRDDPTAVOIHNL 1695  
.1056 KPSVAPK-PDN-----DGSI-----VVLPDGGSPVTVMQNY 1085  
1696 SIIOQTNTVFIAMLEGKGGHYTINISVIYVNSAGAGEKVPMRITMDIKAPARKPKTPPI 1755  
1086 FI-----VWPLRKSRGG-----1098  
1756 YDATCKLLTVSTIIRMPICYYSDDHGPKN-VQULA---TETGAOHGNVWKWDAYFN 1812  
1099 -----QPILLGSPEMDLLEEIIQDLRLQRSLRSRHSQL-----E 1134  
1813 KARPYFTNEGFPNPCTEGTKFS-----GNEEIYIIGADNACMIPONEDKICNGPL 1864  
1135 VPRIYA-----ARFSILPAVPHPNQKY-CGFDPNRGLEPH-----1171  
1865 KPQKYLKFPRATNIMQG----FTDSYSDPV-----KITLGELSERTVEIILSVTLCL 1915  
1172 ---RYVLFLVALVKMQEPTFAASPSPDFQLNDPDOPDIVDG-BEGLIWVGIPVLAV 1225  
1916 SIII LLGTAFAPARIROKEGGTYSPQDAEIIDTKLDOLITVADL---ELKDRELTR 1972  
1226 FII CIVA I LLY-----KKNK-----PDSKRKDSBPRTKCLLNADLAPHHPKDPVEMR 1273  
1973 -----PISKSKSFIQHVELICTNNLKFOBEPSSEL---PKFLQDLSSSTDADLP 2016  
1274 RINFOTPGMLSHPIPITDMAEWERLKAANDSLKIQEYESIDPGOOFTWEHSNLEA--- 1330  
2017 WNAKRNRPNIKYPNNRVKLIAADVSPGDYINASISGYLCPNFEIATOGPLFGTVG 2076  
1331 -NPKPNRYANVIAY-DHSRVILPLEGIMGSDVINANYVDGRRQNAYIATOGPLPETFG 1388  
2077 DFWRMVME TRAKTI VMLTCQFEKGRIRCHQYWEPEDKNKPVTVFGDI VKLMEDVOI-DMT 2135  
1389 DFRWMVQEORSATVMMTRL EESRVC KD QWP-NRGTE TYGI QT-LDDTM ELATFC 1445  
2136 IRDLKIERRHG--DCMTVRQCNFAMPBHGVPENSAPLIHFVKLVRASRAHDTPMPI VHCS 2193  
1446 VRTESLHKNGSSEKREVBHFQTA WPDHGVP EYTPPLA FL RR VK TC NP DP AG PVV H CS 1505  
2194 AGVRTGVFTALDHLTQHINDHFDVIYGVAELSERMCWQNLAQYIFLHQCIIDLIS 2253  
1506 AGVGTGCFFIVADMALERIRTEXTDVYGVHVTTMRSQRMVTMOVEDQYSFIHEALLAEVG 1565  
2254 NKGSNQPICFYNTSYSA LO KMDSL DAME GDV LE WE 2287  
1566 CGNTEVPARSU-YTYIQKLAQVPEGHVTHGWEELE 1598  
  
RESULT 15  
T30938  
receptor tyrosine phosphatase - medicinal leech  
C:Species: Hirudo medicinalis (medicinal leech)  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T30938  
R/Gershon, T.R.; Baker, M.; Nitabach, M.; Wu, P.; Macagno, E.R.  
submitted to the EMBL Data Library, December 1997  
A>Description: Two receptor tyrosine phosphatases of theLAR subfamily are expressed in  
A;Reference number: Z20939  
A;Accession: T30938  
A>Status: preliminary; translated from GS/EMBL/DDBJ  
A:Molecule type: mRNA  
A;Residues: 1-2051 <GER>  
A/Cross-references: UNIPROT: O44328; EMBL: AF017083; NID: g2695654; PID: g2695655; PIDN: AAB9  
C/Genetics:  
A;Gene: LAB2  
C/Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology;  
Cygy

|      |    |      |   |      |
|------|----|------|---|------|
| 1364 | Qy | 1412 | -----VERNSTKVSPQDHRMYTIFIKLANTSVVFKVRASTSA                          | 1465 |
| 960  | Db | 1014 | -----LVVDKLEADRSTFTVRAFTSA  | 1073 |
| 1413 | Qy | 1465 | -----IRPDITILGVFQNYKI   | 1521 |
| 1015 | Db | 1073 | -----GSPWTRNVIKATSSSTALPAPLINIRLQTSVTSMESWDELHGNPLNVLGRIYYSS        | 1119 |
| 1466 | Qy | 1521 | -----TQLRQAQCKEWSBECVYQKIQLYEHAHLTBTETVYGLKKFRWYRQVAASTNAG          | 1567 |
| 1074 | Db | 1119 | -----GPGTSAEVHGLDPSTTYSVRIRARGSDGRFGI                               | 1161 |
| 1522 | Qy | 1567 | -----STKTLPGPPDPPENHVVAVSPFSISISWSEPAVITG                           | 1615 |
| 1120 | Db | 1161 | -----SYSSGRPPRKP  | 1212 |
| 1568 | Qy | 1615 | -----DEFNISFIKSNEENKTIBIKOLEIFTRYSVVITAF                            | 1672 |
| 1162 | Db | 1212 | -----ELMSR  | 1268 |
| 1616 | Qy | 1672 | -----TGKSAEMIVTLESAPKDPDNNMTFQKIPDEVTXFOULTFPPSQPNG                 | 1732 |
| 1213 | Db | 1268 | -----FIRPLSDQMVSRLVK--PAADVLG                                       | 1307 |
| 1673 | Qy | 1732 | -----NIQVYQALVYREDDPTAVQIHNLISIIOKTNTEFVIAMLEGLKGGHTYNI SVVAVNSAGAP | 1346 |
| 1269 | Db | 1307 | -----TQKPRCRGDGL  | 1851 |
| 1733 | Qy | 1791 | -----GQTYSPQDABIIDTKLKLDQLI   | 1384 |
| 1308 | Db | 1346 | -----PMTFTDTPPKGHNVAVL  | 1911 |
| 1792 | Qy | 1851 | -----GRTWYTL  | 1391 |
| 1347 | Db | 1384 | -----WIVAPIC  | 1958 |
| 1852 | Qy | 1911 | -----GQTYSPQDABIIDTKLKLDQLI   | 1441 |
| 1385 | Db | 1391 | -----GQTYSPQDABIIDTKLKLDQLI   | 2016 |
| 1912 | Qy | 1958 | -----GQTYSPQDABIIDTKLKLDQLI   | 1499 |
| 1392 | Db | 1441 | -----GQTYSPQDABIIDTKLKLDQLI   | 2065 |
| 1959 | Qy | 2016 | -----GQTYSPQDABIIDTKLKLDQLI   | 1558 |
| 1442 | Db | 1499 | -----GQTYSPQDABIIDTKLKLDQLI   | 2125 |
| 2017 | Qy | 2065 | -----GQTYSPQDABIIDTKLKLDQLI   | 1615 |
| 1500 | Db | 1558 | -----GQTYSPQDABIIDTKLKLDQLI   | 2182 |
| 2066 | Qy | 2125 | -----GQTYSPQDABIIDTKLKLDQLI   | 1675 |
| 1559 | Db | 1615 | -----GQTYSPQDABIIDTKLKLDQLI   | 2242 |
| 2126 | Qy | 2182 | -----GQTYSPQDABIIDTKLKLDQLI   | 1735 |
| 1616 | Db | 1675 | -----GQTYSPQDABIIDTKLKLDQLI   |      |
| 2183 | Qy | 2242 | -----GQTYSPQDABIIDTKLKLDQLI   |      |
| 1676 | Db | 1735 | -----GQTYSPQDABIIDTKLKLDQLI   |      |
| 2243 | Qy |      | -----GQTYSPQDABIIDTKLKLDQLI   |      |
| 1736 | Db |      | -----GQTYSPQDABIIDTKLKLDQLI   |      |

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 14:42:14 ; Search time 272 Seconds  
(without alignments)  
4313.138 Million cell updates/sec

Title: US-10-673-885-2

Perfect score: 11985

Sequence: 1 MDPLIFLPIFLPIGTSETQVD.....MDSLDMGDEVELEWEETM 2291

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 03.3\*

1: uniprot\_sprot.\*

2: uniprot\_trnbl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match % | Length | DB ID        | Description         |
|------------|--------|---------------|--------|--------------|---------------------|
| 1          | 10536  | 87.9          | 2302   | 088488       | 088488 rattus norv  |
| 2          | 4611.5 | 38.5          | 1086   | 08BY76       | 08BY76 mus musculus |
| 3          | 1323   | 11.0          | 2029   | Q9VIS8       | Q9VIS8 drosophila   |
| 4          | 1319   | 11.0          | 2029   | 1 LAR DROME  | P16621 drosophila   |
| 5          | 1190   | 9.9           | 1597   | Q960M3       | Q960M3 drosophila   |
| 6          | 1188   | 9.9           | 1912   | 1 PTPD HUMAN | P23468 homo sapien  |
| 7          | 1186   | 9.9           | 1889   | 2 Q7QDX2     | Q7QDX2 anopheles g  |
| 8          | 1186   | 9.9           | 1898   | 2 Q9EQ17     | Q9EQ17 mus musculus |
| 9          | 1181   | 9.9           | 1898   | 2 Q64604     | Q64604 r protein-t  |
| 10         | 1172.5 | 9.8           | 1529   | 2 Q6PG86     | Q6PG86 mus musculus |
| 11         | 1168.5 | 9.7           | 1896   | 2 Q9IAJ1     | Q9IAJ1 xenopus lae  |
| 12         | 1162.5 | 9.7           | 1887   | 2 Q9QW67     | Q9QW67 rattus sp.   |
| 13         | 1160   | 9.7           | 1607   | 2 Q8IRSO     | Q8IRSO drosophila   |
| 14         | 1160   | 9.7           | 1767   | 2 Q9W4F5     | Q9W4F5 drosophila   |
| 15         | 1158   | 9.7           | 1894   | 2 Q64487     | Q64487 homo sapien  |
| 16         | 1155.5 | 9.6           | 1897   | 1 PTPF HUMAN | P10586 homo sapien  |
| 17         | 1153   | 9.6           | 1767   | 2 Q24495     | Q24495 drosophila   |
| 18         | 1147.5 | 9.6           | 1948   | 1 PTNS HUMAN | Q13332 homo sapien  |
| 19         | 1144   | 9.5           | 1898   | 2 Q86WS0     | Q86WS0 homo sapien  |
| 20         | 1141.5 | 9.5           | 1556   | 2 Q8IR87     | Q8IR87 drosophila   |
| 21         | 1141.5 | 9.5           | 1631   | 1 PTP1 DROME | P35992 drosophila   |
| 22         | 1141.5 | 9.5           | 1631   | 2 Q86NN9     | Q86NN9 drosophila   |
| 23         | 1141.5 | 9.5           | 1631   | 2 Q9VYV1     | Q9VYV1 drosophila   |
| 24         | 1132   | 9.4           | 1998   | 2 Q8CIW2     | Q8CIW2 mus musculus |
| 25         | 1124.5 | 9.4           | 1788   | 2 Q9IAJ0     | Q9IAJ0 xenopus lae  |
| 26         | 1113   | 9.3           | 1956   | 2 Q6MZFE     | Q6MZFE homo sapien  |
| 27         | 1109.5 | 9.3           | 1904   | 2 Q64699     | Q64699 homo sapien  |
| 28         | 1107   | 9.2           | 1997   | 1 PTPB HUMAN | P23467 homo sapien  |
| 29         | 1086   | 9.1           | 2200   | 1 LAR_CAEEL  | Q9BMM8 caenorhabdi  |
| 30         | 1079   | 9.0           | 1863   | 2 Q64605     | Q64605 rattus norv  |
| 31         | 1044.5 | 8.7           | 5202   | 2 Q6S362     | Q6S362 homo sapien  |

|    |       |     |      |   |            |                         |
|----|-------|-----|------|---|------------|-------------------------|
| 32 | 1033  | 8.6 | 2051 | 2 | 044328     | 044328 hirudo medi      |
| 33 | 987   | 8.2 | 1064 | 2 | Q8SXB2     | Q8SXB2 drosophila       |
| 34 | 902   | 7.5 | 1406 | 2 | Q9W6V5     | Q9W6V5 gallus gall      |
| 35 | 893   | 7.5 | 2222 | 2 | Q7QEG7     | Q7QEG7 anopheles g      |
| 36 | 877.5 | 7.3 | 1102 | 2 | Q8OVN7     | Q8OVN7 mus musculus     |
| 37 | 870   | 7.3 | 2169 | 2 | Q8AV58     | Q8AV58 gallus gall      |
| 38 | 867.5 | 7.2 | 1267 | 2 | Q89945     | Q89945 gallus gall      |
| 39 | 864   | 7.2 | 1254 | 2 | Q8V8V0     | Q8V8V0 mus musculus     |
| 40 | 863   | 7.2 | 1238 | 1 | PTFJ MOUSE | PTFJ MOUSE mus musculus |
| 41 | 861   | 7.2 | 1238 | 2 | Q8K3Q2     | Q8K3Q2 mus musculus     |
| 42 | 860   | 7.2 | 1238 | 2 | Q8CIW9     | Q8CIW9 mus musculus     |
| 43 | 859.5 | 7.2 | 1198 | 2 | Q9JLU0     | Q9JLU0 mus musculus     |
| 44 | 855.5 | 7.1 | 1226 | 2 | Q7TSY7     | Q7TSY7 mus musculus     |
| 45 | 855   | 7.1 | 1367 | 2 | Q20120     | Q20120 caenorhabdi      |

#### ALIGNMENTS

RESULT 1  
O88488 ID O88488 PRELIMINARY; PRT; 2302 AA.  
AC O88488;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Glomerular mesangial cell receptor protein-tyrosine phosphatase  
DE precursor.  
DE Name=PTPRQ;  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Wislar;  
RX MEDLINE=98395110; PubMed=9727007; DOI=10.1074/jbc.273.37.23929;  
RA Wright M.B., Hugo C., Seifert R., Distche C.M., Bowen-Pope D.F.;  
RT "Proliferating and migrating mesangial cells responding to injury  
RT express a novel receptor protein-tyrosine phosphatase in experimental  
RT mesangial proliferative glomerulonephritis.";  
RL J. Biol. Chem. 273:23929-23937(1998).  
DR EMBL; AF063249; AAC34801.1; -;  
DR PIR; T14328; T14328.  
DR HSSP; P10586; 1LAR.  
DR GO; GO:0016787; P:hydrolase activity; IEA.  
DR GO; GO:0004725; P:protein tyrosine phosphatase activity; IEA.  
DR GO; GO:0004872; P:receptor activity; IEA.  
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.  
DR InterPro; IPR003961; FN III.  
DR InterPro; IPR008957; FN III-like.  
DR InterPro; IPR00387; TYR phosphatase.  
DR InterPro; IPR000242; TYR\_PP.  
DR Pfam; PF00041; fn3; 16.  
DR Pfam; PF0102; Y\_phosphatase; 1.  
DR PRINTS; PR00014; FNTYPEIII.  
DR PRINTS; PR00700; PRTYPHPTASE.  
DR SMART; SM00060; FN3; 16.  
DR SMART; SM00194; PTPC; 1.  
DR PROSITE; PS00853; FN3; 14.  
DR PROSITE; PS00383; TYR PHOSPHATASE 1; 1.  
DR PROSITE; PS00056; TYR PHOSPHATASE 2; 1.  
DR PROSITE; PS00055; TYR PHOSPHATASE\_PTP; 1.  
KW Hydrolase; Receptor; Signal.  
FT SIGNAL 1 18 Potential.  
FT CHAIN 19 2302 glomerular mesangial cell receptor  
FT protein-tyrosine phosphatase.  
SQ SEQUENCE 2302 AA; 256822 MW; F0FA703022EB25D5 CRC64;

Query Match 87.9%; Score 10536; DB 2; Length 2302;  
Best Local Similarity 85.9%; Pred. No. 0;  
Matches 1977; Conservative 162; Mismatches 151; Indels 12; Gaps 4;





```

Db 2161 VROCNETGPERGVPENTTFLIHFKVLRSTRADHTTPMVHCSAGVGRGTGVFIALDHLT 2220
QY 2210 QHINDHDFVDIYGLVLAELSRMCMQNLQAQYIFLHQICLDLLSNKSNQIPCFVNSAL 2269
Db 2221 QHINNHDFVDIYGLVLAELSRMCMQNLQAQYIFLHQICLDLLSNKSGHQPVCVFNYSYL 2280
QY 2270 QKMSLDAMEGDGVELEWEETM 2291
Db 2281 QKMSLDAMEGDGVELEWEETM 2302

RESULT 2
Q8BY76 PRELIMINARY; PRT; 1086 AA.
AC Q8BY76;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A630028f16 product:hypothetical fibronectin type III
DE domain/fibronectin type III repeat containing protein, full insert
DE sequence. (fragment).
CN Name=Ptprq;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA The FANTOM Consortium;
RT the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Nishi Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).

```

```

RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akai H., Tanaka Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK041657; BAC31021.1; -.
DR HSSP; P40189; 1BQU.
DR MGD; MGI:1096349; Ptprq.
DR GO; GO:0042472; Primer ear morphogenesis; IMP.
DR InterPro; IPR003962; FnlII subd.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR003534; Royaljelly.
DR Pfam; PF00041; fn3; 9.
DR PRINTS; PRO0014; FNTYPSIII.
DR PRINTS; PRO1366; ROYALJELLY.
DR SMART; SM00060; FN3; 9.
DR PROSITE; PS00853; FN3; 7.
DR Hypothetical protein.
KW Hypothetical protein.
FT NON_TER 1086 1086
SQ SEQUENCE 1086 AA; 120855 MW; 84A18CF2F6857F3C CRC64;

Query Match 38.5%; Score 4611.5; DB 2; Length 1086;
Best Local Similarity 83.2%; Pred. No. 5.5e-233;
Matches 877; Conservative 84; Mismatches 90; Indels 3; Gaps 3;

QY 1 MDPLIIFLLFIQTSETQVDVSNVVPTRYDITISSI-TTYSPTVTRIVTPNVTGPP 59
Db 1 MDPLFFFLSLIGTSESQVDVSGSFDITVYDITLSSISATTSPPVSRILATNVKPGPP 60
QY 60 VFLAGERVGSAGILLSSWNTPPNPGRIISIVKYKEVCPWMTQVTVQVSKPDSLEVL 119
Db 61 VFLAGERVGSAGILLSSWNTPPNPGRIISIVKYKEVCPWMTQVTVQVSKPDSLEVL 120
QY 120 NLNPGTTYEIKVAENSAGIGVSDPFLFQTAESAFKVVNLTVEAYNSAVKLIWYLP 179
Db 121 NLNPGTTYEIKVAENSAGIGVSDPFLFQTAESAFKVVNLTVEAYNSAVNLIWYLP 180
QY 180 QPQKITSPKISVKGARSGIVVKDVSIRVEDILTGLKPECNENSESFLMSTASPSPTLGR 239
Db 181 QPQKITSPKISVKGARSGIVVKDVSIRVEDILTGLKPECNENSESFLMSTASPSPTLGR 240
QY 240 VTPPSRTTHSSSTLTQNEISSVWKBPISFVWTHLRPTTYLFEVSATTEAGVIDSTIVR 299
Db 241 ATPPLETTHLSNLTARNKISSVWKBPISFVWTHLRPTTYLFEVSATTEAGVIDSTIVR 300
QY 300 TPESVPEGPQPCVNTGNITGKFSILWDPPTTVTGKFSYRVLYGSPGRILDNSTKDLKF 359
Db 301 TPESVPEGPQPCVNTGNITGKFSISWDPDPAIVTGKFSYRVLYGFTGAILDNSTKDLRP 360
QY 360 AFTNLTPFTMYDVYIAAETSAGTGPKNISVFTPPDPVPGAFDQLQAEVESTQVITWK 419
Db 361 VFTLTPFTMYDVYIAAETSAGTGPKNISVFTPPDPVPGAFDQLQAEVESTQVITWK 420
QY 420 PRQNGIINOYRVKVLVPTGTGILENTLTGNNEYNIDPKAPEIVNIVPEWGLYEGSAE 479
Db 421 PRQNGIISOYRVKVSLESGVILENTLTGQDEYINNPMPTPEIMMLVDPMGIFGSGE 480
QY 480 MSSDLHSLATFYNSHPDKNFPARNRAEDQTSF-VVYTRNQYITDIAAQLSVVIRRLVP 538
Db 481 MSSDLHSLATFYNSHP-HDFPARKTVEQRFVQVATRNQYMTDIAAHLVSVIRRLVP 539
QY 539 FTEHMISVSAFTIMGEGPPTVLSVTRQOQVPSSIKIINYKNISSSSSILLYWDPPEYNGK 598

```

Db 540 FTEHTISVAFVGMGPGPTVLTVRTREQVPSIQINKNISSSILLYWDPPEYNGK 599  
Qy 599 ITHYTYAMELDTNFAQTITDINSPLITGLKTKYKRWAASTHDSSEIDIVF 658  
Db 600 ITHYTYAMELDTNFAQTITDINSPLITGLKTKYKRWAASTHDSSEIDIVF 659  
Qy 659 RTSEDEPSSPDQVEVIDYADIRLKWSPPKPKNGIITAYEVLKYNIDTLMKNTSTTD 718  
Db 660 RTPEDEPSSPDQVKYTVSPSELSLTWSPPEKPKNGIITAYEVLKYNIDTLMKNTSTTD 719  
Qy 719 IILNRUPHTLKNISVRSYTRGCHGNQVSSLSVRSSTVTPDSAPENITYKNISSEIEL 778  
Db 720 IITLSDLKPYTLNISTQSYTRGCHGNQVSSLSVRSSTVTPDSAPENITYKNISSEIEI 779  
Qy 779 SFLPPSSPNGIITKTYILKRSNGEERTINTSLTQNLKVLKKTQYILVIESASTLKE 838  
Db 780 FFLPPSPNGIITKTYILKRSNGEERTINTSLTQNLKVLKKTQYILVIESASTLKE 839  
Qy 839 GYRSAPISILTBEDAPSPQFVSVKQLSGVTVKLSWQPLPENGILLYTVVWNRSSL 898  
Db 840 GYRSAPISILTBEDAPSPQFVSVKQLSGVTVKLSWQPLPENGILLYTVVWNRSSL 899  
Qy 899 KTNVTETSLSDLDYVVEYSAVYVASTRPGDKTGSNIISFQTPGAPSPDPKDYVA 958  
Db 900 KTNANEVSLSDLDYHADYSAVYVASTRPGDKTGSNIISFQTPGAPSPDPKDYVA 959  
Qy 959 NLSSSSIIILFWPPSPKNGIIQYISVYVYNTSGTFMONTLHETLNDPDMVTSIIDLK 1018  
Db 960 NLSSSSIIILFWPPSPKNGIIQYISVYVYNTSGTFMONTLHETLNDPDMVTSIIDLK 1019  
Qy 1019 TIFSYTFLWTASTSVGNGKSSDIIEVYTDQDI 1052  
Db 1020 AVFSYTYTFLWTASTSVGNGKSSDIIEVYTDQDI 1053  
Qy 1020 AVFSYTYTFLWTASTSVGNGKSSDIIEVYTDQDI 1053  
ID Q9V1S8 PRELIMINARY; PRT; 2029 AA.  
AC Q9V1S8;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)  
DE 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
DE CG10443-PA.  
GN Name-Lar: ORFNames=CG10443;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20196006; PubMed=107311132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yeaman M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle E.C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abrial J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J., Ye J.,  
RA Williams S.M., Woodruff, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195 (2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Javerty T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila  
melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
systematic review.";  
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).  
RN [5]  
RP SEQUENCE FROM N.A.  
RG FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RG FlyBase;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003663; AAF53837.3; -;  
DR HSP; P10596; 1LAR.  
DR FlyBase; FBgn000464; Lar.  
DR GO; GO:0004725; P:protein tyrosine phosphatase activity; IDA.  
DR GO; GO:0007155; P:cell adhesion; IMP.  
DR GO; GO:0008045; P:motor axon guidance; IMP.  
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IDA.  
DR GO; GO:0008360; P:regulation of cell shape; IMP.  
DR InterPro; IPR003962; FNIII subd.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR008957; FN\_III-like.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003598; IG\_c2.



Db 1451 HPIPTSEFANIERLKSNDNQKFSQYESIEPG-QQFTWDSNLSNHNKSNRYANVTAY 1509  
 QY 2031 NNNNRVKLTADSVPSQSDYNASYISGYLCPNEFIATQGLPCTGCDFFRWVWETPAKTL 2090  
 Db 1510 -DHSRVQLPAVEGVGSDYINANYCNGYRKHNAYATQGLQFTFFDFWRCWELKTAIY 1568  
 QY 2091 VMLTQCFEGRIRCHYOVPEDNKPVTVFQDIVITKLMEVDQIDWTIRDLKIERHG--DCM 2148  
 Db 1569 VMTRLEERTRIKCDQYWP--TRGTETYQGFVITETBELATYSIRTFOLCRQGFNDRR 1626  
 QY 2149 TVRQCNTTAMPBGVSPENSAPLTHFVKLVASRAHDTTPMVHCSAGVGRTGVFIADHL 2208  
 Db 1627 EIKQLQFTAMPDGHVDHPAPFLQFLRRCRALTPPESGPFVHCSAGVGRTGCVIIVDSM 1686  
 QY 2209 TQINDHDFVDYIGLVAELRSEKCMQVQLAIFLHCILLDLSNKGNSQPICFVN---- 2265  
 Db 1687 LERMKEHKIIDIYGHVTCRLAQRNVMVQTEDQYIFTHDAILEAI-----ICGVTEVP 1738  
 QY 2266 ----YSALQKM-----DSLDAWE 2279  
 Db 1739 ARNLTHLQKLITEGETISGME 1762

## RESULT 4

LAR\_DROME STANDARD; PRT; 2029 AA.  
 AC P16621;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Protein-tyrosine phosphatase Lar precursor (EC 3.1.3.48) (Protein-tyrosine-phosphate phosphohydrolase) (DLAR).  
 GN Names-Lar;  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90046860; PubMed=2554325;  
 RA Streuli M., Krueger N.X., Tsai A.Y.M., Saito H.;  
 RT "A family of receptor-linked protein tyrosine phosphatases in humans and Drosophila.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:8698-8702(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Canton-S;  
 RX MEDLINE=96178473; PubMed=8598047; DOI=10.1016/S0092-8674(00)81036-3;  
 RA Krueger N.X., van Vactor D., Wan H.I., Gelbart W.M., Goodman C.S., Saito H.;  
 RT "The transmembrane tyrosine phosphatase DLAR controls motor axon guidance in Drosophila.";  
 RL Cell 84:611-622(1996).  
 CC -!- FUNCTION: Possible cell adhesion receptor. It possesses an intrinsic protein tyrosine phosphatase activity (Pirase). It controls motor axon guidance.  
 CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein tyrosine + phosphate.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- TISSUE SPECIFICITY: Selectively expressed in a subset of axons and pioneer neurons in the embryo.  
 CC -!- SIMILARITY: Belongs to the protein-tyrosine phosphatase family. Receptor class 2A subfamily.  
 CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
 CC -!- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC -----  
 CC  
 DR EMBL; M27700; AAC28668.1; -;  
 DR EMBL; U36857; AAC47002.1; -;  
 DR EMBL; U36849; AAC47002.1; JOINED.  
 DR EMBL; U36850; AAC47002.1; JOINED.  
 DR EMBL; U36851; AAC47002.1; JOINED.  
 DR EMBL; U36852; AAC47002.1; JOINED.  
 DR EMBL; U36853; AAC47002.1; JOINED.  
 DR EMBL; U36854; AAC47002.1; JOINED.  
 DR EMBL; U36855; AAC47002.1; JOINED.  
 DR EMBL; U36856; AAC47002.1; JOINED.  
 DR PIR; A36182; TDFELK.  
 DR HSP; P10586; LLAR.  
 DR FlyBase; FBgn0000464; Lar.  
 DR GO; GO:0004725; P:protein-tyrosine-phosphatase activity; IDA.  
 DR GO; GO:0008045; P:motor axon guidance; IMP.  
 DR GO; GO:0006470; P:protein amino acid dephosphorylation; IDA.  
 DR InterPro; IPR003961; FN III.  
 DR InterPro; IPR008957; FN III-like.  
 DR InterPro; IPR003962; FN III subd.  
 DR InterPro; IPR007110; IG c2.  
 DR InterPro; IPR003598; IG c2.  
 DR InterPro; IPR000387; TYR phosphatase.  
 DR InterPro; IPR000242; TYR\_PP.  
 DR Pfam; PF00041; fn3; 9.  
 DR Pfam; PF00047; ig; 3.  
 DR Pfam; PF0102; Y\_phosphatase; 2.  
 DR PRINTS; PR00014; FNTYPEIII.  
 DR PRINTS; PR00700; PRTYPHPTASE.  
 DR SMART; SM00060; FN3; 9.  
 DR SMART; SM00408; IGC2; 3.  
 DR SMART; SM00194; PTPC; 2.  
 DR PROSITE; PS00853; FN3; 9.  
 DR PROSITE; PS00835; IG LIKE; 3.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.  
 DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 2.  
 DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 2.  
 KW Cell adhesion; Glycoprotein; Hydrolase; Immunoglobulin domain; Protein phosphatase; Receptor; Signal; Transmembrane.  
 FT SIGNAL 1 32  
 FT CHAIN 33 2029 Protein-tyrosine phosphatase Lar.  
 FT DOMAIN 33 1377 Extracellular (Potential).  
 FT TRANSMEM 1378 1402 Potential.  
 FT DOMAIN 1403 2029 Cytoplasmic (Potential).  
 FT DOMAIN 36 128 Ig-like C2-type 1.  
 FT DOMAIN 140 224 Ig-like C2-type 2.  
 FT DOMAIN 234 316 Ig-like C2-type 3.  
 FT DOMAIN 322 410 Fibronectin type-III 1.  
 FT DOMAIN 416 508 Fibronectin type-III 2.  
 FT DOMAIN 515 603 Fibronectin type-III 3.  
 FT DOMAIN 610 704 Fibronectin type-III 4.  
 FT DOMAIN 709 807 Fibronectin type-III 5.  
 FT DOMAIN 812 904 Fibronectin type-III 6.  
 FT DOMAIN 909 1002 Fibronectin type-III 7.  
 FT DOMAIN 1006 1099 Fibronectin type-III 8.  
 FT DOMAIN 1101 1203 Fibronectin type-III 9.  
 FT DOMAIN 1492 1738 Protein-tyrosine phosphatase 1.  
 FT DOMAIN 1781 2029 Protein-tyrosine phosphatase 2.  
 FT ACT\_SITE 1670 1670 Phosphocysteine intermediate (By similarity).  
 FT ACT\_SITE 1961 1961 Phosphocysteine intermediate (By similarity).  
 FT DISULFID 57 111 Potential.  
 FT DISULFID 161 209 Potential.  
 FT DISULFID 256 301 Potential.  
 FT CARBOHYD 176 176 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 253 253 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 298 298 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 553 553 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 616 616 N-linked (GlcNAc...) (Potential).

FT CARBOHYD 666 666 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 721 721 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 774 774 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 915 915 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 962 962 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 1183 1183 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 1304 1304 N-linked (GlcNAc...) (Potential).  
SQ SEQUENCE 2029 AA; 229027 MW; 536A0C794D3DC800 CRC64;  
  
Query Match 11.0%; Score 1319; DB 1; Length 2029;  
Best Local Similarity 24.3%; Pred. No. 6.3e-60;  
Matches 530; Conservative 292; Mismatches 762; Indels 600; Gaps 85;  
  
QY 270 VTHLRPTTYTLFEVSAAITAEAGYIDSTIV---RTPEVPEGPQNCVTGNIITGKFSILM 326  
DB 5 MTAARPIAALSLLVLSLT---WTHPTIYDAAPHEIIRKPNQGVGVGASFYCAARG 61  
  
QY 327 DPPTI-----VTGKFSYVELYGPSG-----RILDSNKOLK 358  
DB 62 DPPPISVWRKNGKVGSGTSRYTVLEQPGISILRIEVPVRAGRDADPYECVAENGVDV 121  
  
QY 359 FAFTNLTPETMDVYIAAETSAGTGPKSNISVFTPPDPFGAVFDLQLAEVESTQVRITWK 418  
DB 122 SADATLT-----IYEGDKTPAG-----FPVIT--QGPET---RVIEVGHT-VLMTCK 162  
  
QY 419 KPRQNGIINQYRVKVLVPETGIILENTLLTGNNVEINDMPAPEIWINIVEPMVGLYEGSA 478  
DB 163 AIGNPTP--NIYWIK---NQTKVDMGNPRYSKDGFL-----QIENSREEDQGYECVA 211  
  
QY 479 EMSSDL-HSLATPIY---NSHPDKNEPARNRAEDQTSPPVVTNRNQVITDIAEQLSYVIR 534  
DB 212 ENSMGTEHSAKNLYYKVRVPP-----PTFSRPETISVMLGSLNLSLCIA-----VG 260  
  
QY 535 RLVPFTEHMSVSAFTIMGEGPPTVLSVTRQOVPSIKIINXNLSSSSILLYMDPPY 594  
DB 261 SPMHVKNWKGSEDLT-----PENEMPIGRNVQLINQESA----- 297  
  
QY 595 PNGKITHYTIYAMELDTNRAFOITTDNSFLITGLKXKYTKYKVRVAASDHGESSISEEN 654  
DB 298 -----NYTC-----IAAST-----LGQID 311  
  
QY 655 DIFVTSDEBPSSPODEVIDTAEIRLWS--PPEKPGIILIAEYLYKNIDLYMK 712  
DB 312 SVSVKVKQSLP--TAPTDVQISEVTSATSVLEWSYKGPEDLYVYQYKPKNANQAFSEIS 370  
  
QY 713 NSTSTDIILRNLPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVDPDAPENITVKNIS 772  
DB 371 GIITMYVVRALSPYTEYEFYVIAVNNIGRG--PPSAPATCTTGETWESAPRNVQVRLIS 429  
  
QY 773 SGEIELSFLPPSPNGIHKYTIYLRSGNEERTINT-----TSLTQNIKVLKXYQYII 828  
DB 430 SSTMTVITWEPPTPENGQVYGYKYVYTTNSNQPEASWNSQWDSNELTTSVDTVPHAIYTV 489  
  
QY 829 EYASATLKGEVRSAPISILTBEDAPDSPQDPSVKQLSGVTVKLSWQPLEPENGILLY 888  
DB 490 RVQAYTSMGAGMSTFVQAKAQGVV--SQPSNFRATDIDGETAVTLQWTKPTSSSENIHVY 548  
  
QY 889 TVYVW-----NRSLKTIWVETLSLSDLVNVEYSVAVTASTRFGDCKGTSNIISFOT 943  
DB 549 ELY-WNDTVANQHHRKINSR--AYTLGGLYDPTLYIWLARSQREGAT--TPPIPVRT 605  
  
QY 944 PEGAPSDPPKVVYANLSSSILFWTPP--SKPNGIIQYISVY-----RNTSGTFMQN 996  
DB 606 KQVPGAPPRNITATISSTISLSMLPPVVERSGRIIYKYVFFVVGREDDEATTM-- 663  
  
QY 997 FTILHETNDPDMVTGIIIDKLTFYTFWLTASTSVGNKSSDIIIEYTDQDIPEGF 1056  
DB 664 -TL-----NMT-SIVLDELKRWTEYKIWLVLGTSVGDGPRSHPII--LRTQEDVP-GD 711  
  
QY 1057 VGNLTYESISSATINVSWWPPAQ--NGLVYVYVLSLLOQTPRHVRPPLVTV-----R 1108  
DB 712 PQDVKATPLNSTSIHVSWSKPPLEKORNGIIRGY-----HHAQELRDEGGKFLNE 761

QY 1109 SIYFD-----NLEKYTDYILKITPSTKGFSDTYTAQLYIKTBEDVP----- 1150  
DB 762 PKFVDVDTLEFNVTVTGLQDPTKYSIQVAALTRKGGDR--SAAIVVKTPGCVRVRPTVSLK 820  
  
QY 1151 --ETSPINTFKVLSSTSVLLSWDPVPKPNGAIIISYDL-----TIQGNENYS 1196  
DB 821 INERPIV-----SIELEWERPAQTYGELRGYRLRWGVKQDQALKEMLSGPQ--- 867  
  
QY 1197 FITSNYIILELSPTLYSFPAAARTRKGLGSSILPPTDESVP LAPQMLTINCTS 1256  
DB 868 -MTKCRF---DNLRGVEYEFVAGSNHIGIQETVKIFQTPGTPGGSPSNI----- 917  
  
QY 1257 DFVWLKWSPLPGGIVKVSFKIHEHETDIYKNIISGFKTEAKLVGLEPVSYSIRVS 1316  
DB 918 ----- 917  
  
QY 1317 AFTKVGNGQFNNVFKFTQESVDPVQNMCMATSWOSVLVKWPPPKK--ANGIITQYM 1374  
DB 918 -----TRFQT---PDV-----LCVTNDPPTREHNGIITRYD 946  
  
QY 1375 VTV-----ERNSTKVSQDHYTFIKLANTSVYFKVRASTAGESDESTCHVST 1424  
DB 947 VQFHKKIDHGLGSEENMTL-----RKAVFTNLEENTEIFRVRAVTKQAGPFSDKLIVE 1001  
  
QY 1425 LPETVPSVPTNIAPSDVQSTSATLTWIRPDITILGYFQNYKI--TTQLRAQCKEWESECV 1483  
DB 1002 TERDMGRAPMSIQABATSEQTABIWM--BEVTSRGKLLGYFIYTTWAVEDLDDWQTK--- 1057  
  
QY 1484 EYOKIQLYEAHLTEET--VYGLKFRWYRFQVAASTNAGYGNASNNWISTKTLPGPPDGP 1542  
DB 1058 -----TVGLTESADLVNLEKAQAVAIARFNGLGRLESEKVTYRI---KPEDVP 1105  
  
QY 1543 ENHVAVATSPFSISISWSEPAVITGTCYLDIVKS-----VDNDEFNISFIKNEE--- 1593  
DB 1106 LNLRAHDVSTHSMTLWSPPIRLT--PVNYKISFDAMKVFVDVDSQGSQOTQVPCREILKH 1164  
  
QY 1594 -NKTIEIKDLEIFPTYSVIVITAFGNISAAVYEGESSAEMIYTTLESAPKDPNNWTFQK 1652  
DB 1165 VYKTHITNELSPFTYNNVNSAIPSDYSY-----RPPTKITVYTTQWAAPO--PMVKPDFY 1218  
  
QY 1653 I--PDEVTKFQLTLP--PSOPNGNIQVQALVYREDPTAVQIHNLSIIKQNTFVIALM 1709  
DB 1219 VNGSEI-----LVILPQASEEYGPISHYLVVVPED-----KSNLH----- 1255  
  
QY 1710 EGLKGCHTYNISVYAVNSAGAPKVPIMRITMDIKAPRPK--TKPTPIYDATGKLLVTST 1768  
DB 1256 -----KIPDQPLTDLPLGRNKPERNPAPYIA--AKFPQRSIP 1291  
  
QY 1769 ITIRMPICYSDDHGPIKNVQVLATETGAQHDGNTKWDAYFNKARPYFTNEGFPNPPC 1828  
DB 1292 FTF-----HLGS-----GDDYHN-----FTNR----- 1308  
  
QY 1829 TSGTKFSGNEBIYIIGADNACMI PGNEDKICNGPLKPKKOYLKFKFRATNIMGQ---FTD 1885  
DB 1309 -----KLEREKYRIFRVAVVDTPQKHLYS 1334  
  
QY 1886 SDYS-----DPVTLGEGLSERTVEIILSVTLCSILL 1920  
DB 1335 SPFSFSLDMREAPGGERPHRDPNWPAPPEVSVNKNDE--PEILWVVLPLWSTFIV 1392  
  
QY 1921 GTAFAPARIROKKEGGTYSPODABIIDTKLKLQILITVADLELKDRLTR----- 1972  
DB 1393 STALIVLCVVKRRORPCKT--PDQAAVTRPLMAADLAGPTSPDPVDERLNFQTPGMLS 1450  
  
QY 1973 --PISKKFLOHVBELCTNNLKFQSEFSELPKFIQDLSSTDADLPNRAKKNRFPNKP 2030  
DB 1451 HPPIPISFANHIEKLNDNOKFSQYESIEFG--QOFTWDMNSLNHNSKRYANVTAY 1509  
  
QY 2031 NNNNFKLIADAVPGSDYINASYISGYLCPNEFIATQGPLGTVGDFWVMWETRAKTL 2090  
DB 1510 -DHSRVQLPNEVGVGSYDINANYCDGKYNHAYVATQGLQETFDVFWMCWELKTATI 1568  
  
QY 2091 VMLTQCPEKGRIRCHQYWPEDNKNPVTVFGDIVITKLMEDVQIDWTIRDKIERHG--DCM 2148

Db 1569 VMMTRLEERTRIKCDQWP--TRGTETYQIFVTITETQELATYSIRTFOLCRQGFNDRR 1626  
Qy 2149 TVRQCNFTAPWPHGVSPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVCRGTGFVALDHL 2208  
Db 1627 EIKOLQFTAMPDGHVDPHAPFLQFLRRCAUTPPESGVIIVHCSAGVGRTCYIVIDSM 1686  
Qy 2209 TQIHNDHDFVDYGLVAELSRMCMVQNLQAVIFLHCILDLSSNKGSNQPCIFVN--- 2265  
Db 1687 LERMKHEKILDIYGHVTCURAGNVMVQTEDQYIFIHDAILEAI-----ICGTEVP 1738  
Qy 2266 ----YSALQKM-----DSLAME 2279  
Db 1739 ARNLHTHLQKLITPERGETISGME 1762

## RESULT 5

Q960M3  
ID Q960M3 PRELIMINARY; PRT; 1597 AA.  
AC Q960M3  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE LD45391p.  
GN Name=Lar;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]\_SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY051985; AAK93409.1; -;  
DR HSSP; P10586; 1LAR.  
DR FlyBase; FBgn000464; Lar.  
DR GO; GO:0004725; P:protein tyrosine phosphatase activity; IDA.  
DR GO; GO:0007155; P:cell adhesion; IMP.  
DR GO; GO:0008045; P:motor axon guidance; IMP.  
DR GO; GO:0008470; P:protein amino acid dephosphorylation; IDA.  
DR GO; GO:0008360; P:regulation of cell shape; IMP.  
DR InterPro; IPR003962; FNIII subd.  
DR InterPro; IPR003961; FN III.  
DR InterPro; IPR008957; FN-III-like.  
DR InterPro; IPR000387; TYR\_phosphatase.  
DR InterPro; IPR000242; Tyr\_PP.  
DR Pfam; PF00041; fn3; 8.  
DR PRINTS; PR00014; ENTPEI11.  
DR PRINTS; PR00700; PRTYPHPTASE.  
DR SMART; SM00060; FN3; 8.  
DR SMART; SM00194; PTEC; 2.  
DR PROSITE; PS00853; FN3; 8.  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.  
DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 2.  
DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 2.  
KW Hydrolase.

SQ SEQUENCE 1597 AA; 182017 MW; A8C2B41445636800 CRC64;

Query Match 9.9%; Score 1190; DB 2; Length 1597;  
Best Local Similarity 24.4%; Pred. No. 2 5e-53;  
Matches 420; Conservative 228; Mismatches 555; Indels 518; Gaps 61;  
Qy 684 LKWSPEKPGNIIAYEVLY-----KNIDFLYMKNTSTTDIILRNLRPHLTNINISVR 735  
Db 3 ITWEPETENGQVTKYKVVYTTNSNQPEASWNSQVMDNSLITV--SELTPIAIVTVRQ 60

Qy 736 SYTRFGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGEIELSFLPSSPNIKKYTI 795  
Db 61 AYTSNGAG-PWSTPQVKAQGV--SQPSNFRATDIGETAVLTQWTKPSSHSENIHVHEL 118  
Qy 796 YLKRNGNE---ERINTTSLTONIKVLKKYTOYIIEVSASTLKEGVRSAPISILTEED 852  
Db 119 YWMDTYANQAHHKRISNSEAYT--LDGLYPDPLYIWLAAARSQREGATTPIPVTKQY 176  
Qy 853 APDSPQDPQSVKQLSGVTVKLSWQPP--LEPNGIILYTVYVW-----NRSSLKTINVT 904  
Db 177 VEGAPRRIATATSTTISLSWLPFPPVERSNRIIYKVFVEVGREDDAETATLNM-- 235  
Qy 905 ETSLESLDLDYVVEYSAYVTASTREFGDKGTGNSIIISFQTPEGAPDPKDVVYANLSSSS 964  
Db 236 -TSIVLDELKRWTEYKIWLACTSVGDGPRSHPII-LRTQEDVPGD-PQDVKATPLNSTS 292  
Qy 965 IILFWTPPSKP--NGIIQYISVYVYRNTSGTFMQNFTLHLELTFNDFNMVSTIIDLKLTFS 1022  
Db 293 IHVSWKPPLEKDRNGIIRGYHIH-----AQEL----- 319  
Qy 1023 YVTFWLTASTSVGNGKSSDIIIEVYTDODIPEGFVGN-LTYESISSTAINSVVPPAQN 1081  
Db 320 -----RDEGKGFLENEPFKFDVVDTLFENVTGLQPD--- 349  
Qy 1082 GLVFFYVSVLILQOTPRHVRPPLVTVYERSIYFONLEKYTDYILKITPSTEGFSDTYTAOL 1141  
Db 350 -----TKYSIQVAALTRKGDGR-SAAI 371  
Qy 1142 YIKTEEDVP-----ETSPINTPNLSSTSVLLSWDPVPKPNGAILSYDL----- 1186  
Db 372 VVKTPGVVPVPTVSLKIMEREPIV-----SIELEWERPAQTYGELRGYRLRWGVK 422  
Qy 1187 -----TLQGNENYSITDNYIILBELSPFTLYSFFAAARTKGLGSPSILFFYDTE 1239  
Db 423 DOALKEMLSGEQ-----MTKKRP---DNLEKGVYEFRAVAGNSHIGIGQETVKILQPTDE 474  
Qy 1240 SVPLAPPQNLTILNCTSDTFVWLKWSPLPGGIVKVSFKIHEHETDTIYYKNISGFKTE 1299  
Db 475 GTFGPPPSNIT----- 485  
Qy 1300 AKLVGLEVPVSTYIRVSAPFTKVGNGQNSVNVKFTQBSVPPDVVQNMCMATSWOSVLVK 1359  
Db 486 -----IRFQT---PDV-----LCVT 497  
Qy 1360 WDPKKK--ANGIIQYVMYTV-----ERNSTKVSPQDHWYTFIKLLANTSVEVKYR 1407  
Db 498 WDPPTREHRNGIITRYDVQFHKIDHGLSGERNMTL-----RKAVFTNLEENTEIFRVR 552  
Qy 1408 ASTSAGEGDESTCHVSTLTPETVPSTNIAFSDVQSTSATLTWIRPDTILGYFQNYKI-T 1466  
Db 553 AYTQAGAGFPFSDKLIVETERDMGRAPMSIQAEATSEQTAEIWW-EPVTSRGKLLGKIFY 611  
Qy 1467 TOLRAQCKEWESECEVYQKIYLYEAHLTEET-VYGLKKPRWYRFOVAATNAGYNA 1525  
Db 612 TWTAVEDLDDWQTK-----TVGLTESADLVNLEKPAQAVAIARFKNGLGLRL 659  
Qy 1526 SNWISTKTLPGPPDGPENNVHVATSPFSISISWSEPAVITGPTCYLLDVKS---VND 1581  
Db 660 SEKVTVRI---KPEDVPLNLRADHVSTHMTLSWSPPIRLT-FVNYKISFDMKVPVDSQ 715  
Qy 1582 EFNISFIKSNEB-----NKTIEIKDLIEFTYISVVITAFITGNISAAVVEGSSAEMIVT 1635  
Db 716 GFSQTIQVPKREIILKHVYKTHINELSPFTTYNVNVSAPSDYSV-----RPPKIITVT 770  
Qy 1636 TLESAPKDPNNMTFQKI--PDEVTKFQLTFILP-PSQPNGINQVYQALVYREDDTAVOI 1692  
Db 771 TQMAAPQ-PWVKPDPFVGVNGBEI---LVILPOASEYGPISHYLVVVPED----- 818  
Qy 1693 HNLIIQKNTNTVIAMLEGLKGGHTYINSVAVNSAGAPKVPMTIMDIKAPAREK-TK 1751  
Db 819 -----KSNLH-----KIPDQFLTDLLPGRNKRPER 843  
Qy 1752 PTPYIDATGKLLVSTTTITIRMPICYSDDHGPIKNVQVLATETGAQHDGNVTWKYDAYF 1811

Db 844 PNPYIA-AKFPORSIPFP-----HLGS-----GDVH 871  
Qy 1812 NKARPYTFNPGFPNPPCTGCKTSGNNEIYIIGADNACMPGNEKICNGPLKPKQYL 1871  
Db 872 N-----FTNR-----KLEREKR 885  
Qy 1872 PKFRATNMQ-----FTDSYVS-----DPVKTGEGLSERT 1903  
Db 886 IFVRVAVDTTPQKHLTYSPFSEFLSLDMREAPPGERPHRDPNPWPAEPEVSVNRNDE-- 943  
Qy 1904 VEILSVTLCLISILLGTAFAPARIRKQKGGTVSPDARIIDTKLKLQDQITVADL 1963  
Db 944 PELWVLPWLVSTFIVSTALIVLCVVRKRRQCKT--PDQAAVTRPLMAADLGAGPTPS 1001  
Qy 1964 ELKDERLTR-----PISKKSFLQHVVEELCTNNLKFQBEFSELPKFLQDLSTDA 2013  
Db 1002 DPVDMRLNFQTPGMISHPPPIPISEFANHLERLKSNDNQKFSQYESIERP-QOFTWDNS 1060  
Qy 2014 DLPWNAKRNPNKYPNNNRVKLJADASVPGSDYINASYISGYICPNFEIATQGLPG 2073  
Db 1061 NLEHNSKRYANVTAY-DHSRVQLPAVEGVGSDYINANYCDYRKNAYAVATQGPLOE 1119  
Qy 2074 TVGDFRMWVETRAKTLVMTQCFEKGRIKCHOYWPEDNKPVTVPFGDIVITKLMEDVQID 2133  
Db 1120 TVDFWRMCWELKTATIVMTRLEERTRIKCDQWP--TRGTETYGOIFVTITETQELAT 1177  
Qy 2134 WTIRDLKIEHGG--DCWTRQCNFTAPWPHGVGPENSAPLIHFKVLVRASRAHDTTPMIVH 2191  
Db 1178 YSIRTQLCQGFNDREIKQLFTAMPDHGVDPHAPFLQFRLCRALTPPESGPVIVH 1237  
Qy 2192 CSAGVGTGVFIADHLTHQINDHDFVDIYGLVLAERLSEKVMQVLAQYIFLHCILDL 2251  
Db 1238 CSAGVGTGVFIADHLTHQINDHDFVDIYGLVLAERLSEKVMQVLAQYIFLHCILDL 2251  
Qy 2252 LSNKGSNQPCFVN-----YSALQKM-----DSLDAME 2279  
Db 1298 I-----ICGVTEVPARNLTHLQKLTETGETISGME 1330

## RESULT 6

PTPD\_HUMAN STANDARD; PRT; 1912 AA.  
AC P23468;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Protein-tyrosine phosphatase delta precursor (EC 3.1.3.48) (R-PTP-  
delta).  
GN Names:PTPRD;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE FROM N.A., AND MUTAGENESIS OF ARG-1178.  
RX MEDLINE=95204468; PubMed=7896816; DOI=10.1074/jbc.270.12.6722;  
RA Pulido R., Krueger N.X., Serra-Pages C., Saito H., Streuli M.;  
RT "Molecular characterization of the human transmembrane protein-  
tyrosine phosphatase delta. Evidence for tissue-specific expression of  
RT alternative human transmembrane protein-tyrosine phosphatase delta  
RT isoforms.";  
RL J. Biol. Chem. 270:6722-6728 (1995).  
RN [2]  
RP SEQUENCE OF 390-1912 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=91006018; PubMed=2170109;  
RA Krueger N.X., Streuli M., Saito H.;  
RT "Structural diversity and evolution of human receptor-like protein  
RT tyrosine phosphatases";  
RL ENBO J. 9:3241-3252 (1990).  
RN [3]  
RP INTERACTIONS WITH PPPIA1; PPPIA2 AND PPPIA3.

RX MEDLINE=98288299; PubMed=9624153; DOI=10.1074/jbc.273.25.15611;  
RA Serra-Pages C., Medley Q.G., Tang M., Hart A., Streuli M.;  
RT "Lipins, a family of LAR transmembrane protein-tyrosine phosphatase-  
interacting proteins.";  
RL J. Biol. Chem. 273:15611-15620 (1998).  
RN [4]  
RP SPLICE ISOFORM(S) THAT ARE POTENTIAL NMD TARGET(S).  
RX PubMed=14759258; DOI=10.1186/gb-2004-5-2-r8;  
RA Hillman R.T., Green R.E., Brenner S.E.;  
RT "An unappreciated role for RNA surveillance.";  
RL Genome Biol. 5:RESEARCH008.1-RESEARCH008.16(2004).  
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein  
CC tyrosine + phosphate.  
CC -1- SUBUNIT: Interacts with PPPIA1, PPPIA2 and PPPIA3.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Comment=Additional isoforms seem to exist;  
CC Name=1;  
CC IsoId=P23468-1; Sequences=Displayed;  
CC Notes=May be produced at very low levels due to a premature stop  
CC codon in the mRNA, leading to nonsense-mediated mRNA decay;  
CC Name=2; Synonyms=Kidney;  
CC IsoId=P23468-2; Sequence=VSP\_005147, VSP\_005148, VSP\_005149;  
CC Name=3; Synonyms=Fetal brain;  
CC IsoId=P23468-3; Sequence=VSP\_005150;  
CC -1- PTM: A cleavage occurs that separates the extracellular domain  
CC from the transmembrane segment.  
CC -1- SIMILARITY: Belongs to the protein-tyrosine phosphatase family.  
CC Receptor class 2A subfamily.  
CC -1- SIMILARITY: Contains 8 fibronectin type III domains.  
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
CC -1- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L38929; AAC41749.1; -;  
CC EMBL; X54133; CAA38068.1; -;  
CC PIR; A56178; A56178.  
CC HSP; P10586; ILAR.  
CC Genes; HGNC:9668; PTPRD.  
CC MIM; 601598; -;  
CC GO; GO:0005887; C:integral to plasma membrane; TAS.  
CC GO; GO:0005001; F:transmembrane receptor protein tyrosine pho. . .; TAS.  
CC GO; GO:0006470; P:protein amino acid dephosphorylation; TAS.  
CC GO; GO:0007185; P:transmembrane receptor protein tyrosine pho. . .; TAS.  
CC InterPro; IPR003961; FN\_III.  
CC InterPro; IPR008957; FN\_III-like.  
CC InterPro; IPR003962; FNIII subd.  
CC InterPro; IPR007110; IG-like.  
CC InterPro; IPR003598; IG\_c2.  
CC InterPro; IPR000387; TYR phosphatase.  
CC InterPro; IPR000242; TYR\_PP.  
CC Pfam; PF00041; fn3; 8.  
CC Pfam; PF00047; ig; 3.  
CC Pfam; PF0102; Y\_phosphatase; 2.  
CC PRINTS; PR0014; FNTYPEIII.  
CC PRINTS; PR00700; PRTYPHPTASE.  
CC SMART; SM00060; FN3; 8.  
CC SMART; SM00408; IGC2; 3.  
CC SMART; SM00194; PTEC; 2.  
CC PROSITE; PS50853; FN3; 8.  
CC PROSITE; PS50835; IG\_LIKE; 3.  
CC PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.  
CC PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 2.  
CC PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 2.  
CC Alternative splicing; Glycoprotein; Hydrolase; Immunoglobulin domain;  
KW



|    |   |  |
|----|---|--|
| KW | Protein phosphatase; Receptor; Repeat; Signal; Transmembrane. |  |
| FT | SIGNAL  | 1 20 Potential.  |
| FT | CHAIN   | 21 192 Protein-tyrosine phosphatase delta.                       |
| FT | DOMAIN  | 21 1265 Extracellular (Potential).                               |
| FT | TRANSMEM  | 1266 1290 Potential.   |
| FT | DOMAIN  | 1291 1912 Cytoplasmic (Potential).                               |
| FT | DOMAIN  | 24 114 Ig-like C2-type 1.  |
| FT | DOMAIN  | 126 224 Ig-like C2-type 2.                                       |
| FT | DOMAIN  | 236 318 Ig-like C2-type 3.                                       |
| FT | DOMAIN  | 323 411 Fibronectin type-III 1.                                  |
| FT | DOMAIN  | 417 511 Fibronectin type-III 2.                                  |
| FT | DOMAIN  | 516 604 Fibronectin type-III 3.                                  |
| FT | DOMAIN  | 609 706 Fibronectin type-III 4.                                  |
| FT | DOMAIN  | 711 819 Fibronectin type-III 5.                                  |
| FT | DOMAIN  | 824 913 Fibronectin type-III 6.                                  |
| FT | DOMAIN  | 918 1013 Fibronectin type-III 7.                                 |
| FT | DOMAIN  | 1017 1103 Fibronectin type-III 8.                                |
| FT | DOMAIN  | 1375 1618 Protein-tyrosine phosphatase 1.                        |
| FT | DOMAIN  | 1619 1912 Protein-tyrosine phosphatase 2.                        |
| FT | ACT_SITE  | 1553 Phosphocysteine intermediate (By similarity).               |
| FT | ACT_SITE  | 1844 Phosphocysteine intermediate (By similarity).               |
| FT | SITE  | 1175 1178 Cleavage (Potential).                                  |
| FT | CARBOHYD  | 254 254 N-linked (GlcNAc. .) (Potential).                        |
| FT | CARBOHYD  | 299 299 N-linked (GlcNAc. .) (Potential).                        |
| FT | CARBOHYD  | 724 724 N-linked (GlcNAc. .) (Potential).                        |
| FT | CARBOHYD  | 832 832 N-linked (GlcNAc. .) (Potential).                        |
| FT | VARSPLIC  | 181 189 Missing (in isoform 2).                                  |
| FT | VARSPLIC  | 226 229 Missing (in isoform 2).                                  |
| FT | VARSPLIC  | 775 783 Missing (in isoform 2).                                  |
| FT | VARSPLIC  | 609 1137 Missing (in isoform 3).                                 |
| FT | MUTAGEN   | 1178 1178 R->A: 2.5-fold reduction in cleavage.                  |
| FT | SEQUENCE  | 1912 AA; 214759 MW; 3AE8CBDD32182E26 CRC64;                      |
| QY | Query Match   | 9.9%; Score 1188; DB 1; Length 1912;                             |
| DB | Best Local Similarity   | 22.9%; Pred. No. 4.2e-53;  |
| QY | Matches   | 487; Conservative 285; Mismatches 724; Indels 630; Gaps 76;      |
| QY | 296   | TIVRTPESEPGP-PONCVTGNITGKSFSLMDP-PTIVTGKFSVRV-----ELYCP 345      |
| DB | 18  | TDAETPPRTRIPVDQVSGGVASFICQATGDRPKIVNNKGGKVSQRFVIEFDG 77          |
| QY | 346   | SGRIIDNSTKDLKFAFTNLTPF-----TWYDVYIAAETSAGTGPKSNISVFTPPDVGCA 399  |
| DB | 78  | SGSVL-----RIQPLRTPRDEAIYEC-VASNNVGEISVSTRILVLRDQIPRG 124         |
| QY | 400   | --VFDL--QLAEVESTQV-----RITWKKPRQP-----NGIINQYRVKULV 436          |
| DB | 125   | FPTIDMGPKLVKVERTRTATMLCAASGNPDPEITWFKDLPDVTNSNNGRIKQLRSE---- 181 |
| QY | 437   | PETGIILENTLLTGNNYEINDPAPRIVNIPEWVGLYEGSDEMSSDLHSLA---TFIYN 493   |
| DB | 182   | -----SIGTPIRG-----ALQIEQSESDQKIECVATNSAGTRYSAPANLYRE 227         |
| QY | 494   | SHPKNFPARNRAEDQTSPTVTRNQYITDIAAEQLSVYIRRLVPPTEHMSVSAPFTMG 553    |
| DB | 228   | LREVRVPRPSIPPTNHEIMPGSVNITCVA-----VGSMPMPVKVMGLGAEIDLTPED 280    |
| QY | 554   | EGP--PTVLVSRTRQ-----VPSIKIINYNI-----SSSSIL 586                   |
| DB | 281   | DMPIGRNVLELNDVROSANYTCVAMSTLGVIEAIAQITVKALPKPGTPVWTESTATSIT 340  |
| QY | 587   | LYWDP--PEYNGKITHYIYAMELDTNEAFQITTD-----NSFLITGLKKYTKYKRV 640     |
| DB | 341   | LWDSGNPE-----PVSYYIIQHKPKNSEELYK--EIDGVATTRYVAGLSPSYDVEFRVV 394  |
| QY | 641   | ASTHDGESSLSEENDIFVRTSEDEPPSPQDVEIDVTADEIRLKWSPPEKPNGLIIAYE 700   |

|    |      |   |
|----|------|---|
| DB | 395  | AVNNIGRGPSE--PVLQTSEQAPSSAPRDVQARMLSSTTILVQWKEPEEPNGOIQYR 452     |
| QY | 701  | VLYKNIDTLYMKNVTSTTDI-----LLRNLRPHLYNLSVRSYTRFGHGNQVSSLLSVRT 754   |
| DB | 453  | VYTTWDPQHVNNWNNKHNVAQSQITIGNLVPKQTSYVKVLAFTSIGDG-PUSSOIQVIT 511   |
| QY | 755  | STVTPDSAPENITYKNISSGEIELSFLPPSSNGIIKKYTIYLRSGNNEERTINTTSIT 814    |
| DB | 512  | QTGVP-GQPLNPKAEPESETSIILSWTPPRS-DTIANYELVYKDGEGEQRIITEPGNS 569    |
| QY | 815  | QNIKVLKXYQYIIIVSASTLKGEVRSAPISEILTEADAPDPPQDFSVKQLSGVTVKLS 874    |
| DB | 570  | YRLQGLKPNLSYIFRLAARSPOGLGASTAEISARTWQSKSPAPPQDISCTSPSSILVS 629    |
| QY | 875  | WQPLEPENGIIILYYIVVWNRSSLKTINVTESLSLSLDLYNVVEISAYVTASTRGDKT 934    |
| DB | 630  | WQPP-----   |
| QY | 935  | GSNIISFQTPEGAPDPKDVVYANLSSSIILFWTPPSKPNGLIIQYYSVYVYRNTSGTFM 994   |
| DB | 634  | -----PVEKQNGIITEYSIKYTAVDG--- 654                                 |
| QY | 995  | QNFTHLHNTDNDNMVTSTIDKLTIFSYTYFWLTAFTSVGNKSSDIIEVTTODIPE 1054      |
| DB | 655  | EDDKPHEILG-IPSDTKYLLQLEKWTETRYITVTAHTDVGPGPESLSVL-IRTNEDVPS 712   |
| QY | 1055 | GFVGNLTYESISSTAINVSW---VPPAQPNGLVFFYVVSILLQOTPRHVRPPLV----- 1104  |
| DB | 713  | GPGRKVEAVNNTSVKVSWSRSPVKNQHQGRGQVHYVRMENGEPKGQPMKDKVMLAD 772      |
| QY | 1105 | -----TYERSIYPDNLEKYTDYILKITPTEKGFSDTYTAQLYIKTEEDVPETSPIN 1157     |
| DB | 773  | AQWEDDDTTEHDMIIISGLQPETSYSLTYTAYTKGDGARSKEPL-VSTTGA VPKRPLV- 830  |
| QY | 1158 | TFKNLSSTSVLSWDPVPKPGNGLIISYDLTLQGNENYSFITSDNYIILBELSPFTLYSF 1217  |
| DB | 831  | -INHTQMTALIQWHPVDVTFGLQGYRLKF-----                                |
| QY | 1218 | FAAATRKGLGPSSLIFFYTDSDSVPLAPPQNLTINCTSDFWLKWSPSPPLGGIVKVYS 1277   |
| DB | 861  | -----GRKDEPPLTLFSEKE-----   |
| QY | 1278 | FKIHEHETDTIYKKNISGFKTEAKLVGLEPVSYSIRVSFAFTKVGNGQFNSVVKFTQER 1337  |
| DB | 878  | -----DHFTATDIHKAS-----YVFLSARNKVGFGEM--VKEISIEP 915               |
| QY | 1338 | SVP-DVQNMOCMATSWOSVLVKWDPKKA--NGIITQY-----MVTVERNSTKV 1384        |
| DB | 916  | EVPTGPPQNLHSEGTSTSVQLSQWPPVLAERNGIITKYTLRYDRINIPLLPMEQ---LI 972   |
| QY | 1385 | SPQDEHYTFIKLLANTSVYFKVRASTSGEGDES-TCHVSTLPETVPSPV-PNTIAFSDVQ 1442 |
| DB | 973  | VADTWTMLTGLKPDVTVKVRHTSKGPGYSVQPRLLP--VDQVFAKNFHVKAWM 1030        |
| QY | 1443 | STSATLTWIRPDITILGYFONYKIITTLQRAQCKEWESECEYQKIQYLYEAHLTEETVY 1502  |
| DB | 1031 | KTSVLLSWEIP-----ENYSNAMPFKIL----YDDGKWEE-----VDGRATOKLIV 1073     |
| QY | 1503 | GLKKFRWYRFQVAASTNAG--YGNASNWISTKTLPGPDGPPENVHVATSPSISISWS 1560    |
| DB | 1074 | NLKPEKSYFVL---TNRGSAGGLQHRVTAKTAP-----DVLATKP----- 1112           |
| QY | 1561 | EPAVITGPTCYLLDVKSVNDDEFNISFKTSNENETKIEIKOLEIFTRYSVVITAFGTNIS 1620 |
| DB | 1113 | -----API-----   |
| QY | 1621 | AAVVEGKSAEMIVTTLESAPKDPNNMTFKIPDEVTKFQLTFLPPSPQNGNIQVYQAL 1680    |
| DB | 1116 | -----GKNLDGMIIV--QLPEVPAN-----ENIKGYIIIVPLKKSRGK-----P 1154       |
| QY | 1681 | VYREDDDPTAVQIHNIL--SIIQKNTFVIAMLBGLKGHTYNTISVAVNSAGAGPKVPMRI 1738 |
| DB | 1155 | IKPWESPDEMDLDELKEISRK-----RSIRYGREVELK- 1189                      |

```

QY 1739 TWDIKAPARKTPTIYDATGKLVSTIIRMPICYSDHGHGPIKQVQVLATETGAQ 1798
DB 1190 -----PYIAAHFDVLPTEFTL-----GDD-----K 1209
QY 1799 HDGNVTWKYDAYENKARPVFTNEGFPNPPCTEGTKFSGNEEYIIGADNACMIPGNEDEK 1858
DB 1210 HVG-----GFTN-----KQLOSQGEVVPFVLA-----VMEHAEK 1239
QY 1859 ICGNPLPKKQYLKFKFRATNIMQFTDSDYDPVKTLG-----EGLSERTVEIIL 1908
DB 1240 M-----YATSPYSDPVVSMOLDPOPIITDEBEGLI-----WVV 1271
QY 1909 SVTLCLLSIILLCTAIFAPARIRKQKEGTVSPQDAIID-----TKLKLDQILITVADLE 1964
DB 1272 GPVLAVVFVFIICVIAILLKRAESDSRKSSIPNNKEIPSHHTDPVLELRL-----NFQ 1327
QY 1965 LKDERLTRPISKKSFLQHVVELCTNNLKPQEBFSELPKFLQDLSLTDADLPWNAKRF 2024
DB 1328 TPGMASHHPPIPILELADHIERKANLNFQSBYESIDFG-QQFTWEHENLEVNKPKNY 1386
QY 2025 PNKYPNNNRVKLIADASVPGSDYINASYIGYLCNPFIAIATQGLPCTVGDPMRWWE 2084
DB 1387 ANVIAY-DHSRVLLSAIEGIPGSDYVANYIDYRKQNAVIAIATQGLPETFGDFRMIWE 1445
QY 2085 TRAKTLVMTQCFEKGRIHQWPNEDKPVTVFGDIVITKLMEDVOI-DWTIRDLKIER 2143
DB 1446 QRSATVMMTKLEERSRVCDOYWP--SRGTETHGLVQVT-LIDTVELATYCVRTPALYK 1502
QY 2144 HG--DCMVTROCNFTAMPBHGVPENSAPLIHFVKLVRAASHDTTMMIVHCSAGVGRTCV 2201
DB 1503 NGSSKREVRQFTAMPDHGVPEHPTFLAFURVKTCNPPDAGPMVHVHCSAGVGRTCG 1562
QY 2202 FIALDLHTQINDHDPVDIYGLVAELSRMCMQVLAQYIFLHQILDLNKGSGNQPI 2261
DB 1563 FIVIDAMLERIKHEKTVDIYGHVTLMAQRNVMVQTEDQYIFHDALLEAVTCGTEVPA 1622
QY 2262 CFVNTSALQKMSLDAMEGDVELEWE 2287
DB 1623 RNL-YAYIQKLTQIETGENVTGMELE 1647

RESULT 7
Q700X2
ID Q700X2 PRELIMINARY; PRT; 1889 AA.
AC Q700X2;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE EbiPel103 (Fragment).
GN Name=ebiG6103; ORFNames=ENSANGG000000004649;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera, Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008980; EAA14539.1; -.
DR HSP; P18052; IPI5.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004725; F:protein tyrosine phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR003962; F:III subd.
DR InterPro; IPR003961; F:III.
DR InterPro; IPR008957; F:III-like.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR000387; TYR_phosphatase.

```

```

DR InterPro; IPR000242; TYR_PP.
DR Pfam; PF00041; fn3; 8.
DR Pfam; PF0102; Y_phosphatase; 2.
DR PRINTS; PRO0014; PRTYPEI11.
DR PRINTS; PRO0700; PRTYPHPHTASE.
DR PROSITE; PS00853; FN3; 8.
DR PROSITE; PS00835; IG_LIKE; 3.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR PROSITE; PS00506; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS00555; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase.
FT NON_TER 1 1889
FT SEQUENCE 1889 AA; 213021 MW; 531EFEBD4C13BD86 CRC64;
Query Match 9.9%; Score 1186; DB 2; Length 1889;
Best Local Similarity 23.1%; Pred. No. 5.3e-53;
Matches 488; Conservative 282; Mismatches 726; Indels 618; Gaps 78;
QY 301 PESVPEGPPONCVGNITGKSFSLWDP-PTVTGKFSYRVLYGSGRILD-NSTKDLK 358
DB 2 PEIIKKPVNQGVGVGATFFCAARGDQPTVIMRKNSKKIMMTOSTRYTYVDANGVSMRL 61
QY 359 P---AFTNLTPFTMYDVVIAAETSAG--TGPKSNISVP---TPDVP----- 397
DB 62 IEPVRAGRDDADYE-----CVAENGVDGAVSAEALTVYEGDKTPAGFPVVKQPTIRVI 116
QY 398 --GAVFDLQAEVESTQVRITWKPKRQPNGIINQYRVKLVLPETGIIILENTLLTGNNYI 455
DB 117 EGTHTAVMQCKATGSPPKIYWLK-----DMKRVDMTNPRYSINSGSLQIDNSE-- 166
QY 456 NDPMAPEIIVNIVPMVGVYEGSAEWS--SDLHSLATFIY-----NSHPDK----- 498
DB 167 -----ESDMGKYECVAENSIGTEHTKPTPLYVKVRVPPTFSRPPEPVYEVML 214
QY 499 --NPPARNRAEDQTSVPVTRNOYITDAAEQLSVYIRLVLPPTHEMISVSAFTINGEGP 556
DB 215 GANLTLTCAVAGSPMPSVKWRKGVDDLTPENDVPVGRNVLELDIRVSTN-YTCAQSS 273
QY 557 PTVL---SVTRQQVPSIKIINYKNISSSILLYWDPPPEYNGKITHYTIYAMELDTNR 613
DB 274 LGVIEATSLVKVQSLPAAPTDTISETVATQVRLEW---SYKGPEDLQY--YVIQYKPN 328
QY 614 AFOITTDINSFLIT-----GLKYYKYKMRVAASHTDGESSLSSENDIFVRTSEDEPS 667
DB 329 ANQASIVEISGIITFMFYVVRTLSPTYEYFYVIAVNNIGRPPS--LPATTTTGETEMES 386
QY 668 SPQDVEVDVDAEIRLAKWSPEKENGIIIAEVLY-----KNIDTLYMKNTSTTDI 719
DB 387 APRNIEVKPLSSSTWLTITWEPETPNQVTKYKYVYTTNPNQPEASWDSQMETNDMTT-- 444
QY 720 ILRNLRPHTLYNISVRSYTRFGHGNQVSSLSVRSRTSETPDSAPENITYKNISSGRIELS 779
DB 445 -ISELTPHAIYIRVOAFTSKGAG-PLSNPVQKAAQGV-PQPSNFRATDVGETAVTLQ 501
QY 780 FLPPSSPNGIKKYTYILKRGNGNE---ERTINTSLTONIKVLKKYQYIIIEVSASTLK 836
DB 502 WSRPTHSGENIVHVELYWNDDTVANEQHQRIPTNTETYT--LSSLYPDTLYYFWLSARSQR 559
QY 837 GEGVSAISILTEEDAPDSPPQDSVKQLSGVTYKLSWQPP--LEPNCIILYYTYV--- 891
DB 560 GEGATTPPIPVRTKQYVFCAPPNRYVTEATSPITINVSMLPVPVRSNGAIYVYKVPFVE 619
QY 892 YMNRSLSLTI-NVTETSLSDLDYVVEYSAVVYASTFRGDCGKTSNIIISFOTPEGAPSD 950
DB 620 VGRSDEATVTTLNSTSIVLDELKWTYKIKVLWLAGTSVGDGPR-SYPYVTRHEDVPGD 678
QY 951 PPKQVYYANLSSSSIIILFWTPPSKE--NGIIQYYSVYRTNSGTGMQNFTHLHFNDFDN 1008
DB 679 -PQDVKASPVNSTTIIVTWKPKKRNNGIIRGYHIH-----VQEMKEE--- 721
QY 1009 MTVSTIIIDKLITFSYTTFWLTASTSVGNGNKSDDIIIEVTVQDIPGFGVGNLTYSISST 1068

```

Db 722 -----GKG-----LLNEPMKFDVVDGLEYNVT----- 743  
Qy 1069 AINVSVPAPQNGNLVYVYVSLILQTPRHVRPLVTVYERSIVFDNLEKVTYDILKITPS 1128  
Db 744 -----GLQDPDKYSVQVAAL 758  
Qy 1129 TBKGFSDTYTAQIYIKTEBDVPETSPINTFNKLS-----STSVLLSWDPVPKNGAI----- 1181  
Db 759 TRKGDGDR-SAPISVKTGGVP-IRETV-TLAILERDPTVSIELERPRQTYGELRGYR 815  
Qy 1182 ISVDLTQGPNNYSFITSNDVILILELSPFTLYSPFAARTRKGLGPSIIILFFTYDES 1241  
Db 816 VRWGREQALNEILQGTQLNKRINDLRGVEYEFVRAGMNHIGIGQEAHVHLQTPGS 875  
Qy 1242 PLAPPQNLTLINCTSPFWLWKSPSPPLPGGIVKVSFKIHEHET-----DRIYKNSI- 1294  
Db 876 PTGPPTGIAVRFPDVCITWEP-----PTREHNGOITRYDVOFHKKIDH 922  
Qy 1295 GFKTE-----AKLVGLEPVSYSIRVSAFTKVGNGNOFSNVVKTQTESVDPVQNMQ 1347  
Db 923 GLGTERNTTVKRAVFTNLDESTEYIVRVAYTKQAG-PFSEKVIATERDMGRAPFSVQ 981  
Qy 1348 CMATSQSVLVKWDPPKANGIITQVMTVERNSTKVSQDHYTFIKLLANTSVYFKVR 1407  
Db 982 AVATSEQTVEVWEP----- 996  
Qy 1408 ASTSAGEBDESHVSTLPTVSPVPTNIAPSDVQSTSATLTWIRPDITILGYFQNYKI-T 1466  
Db 997 -----VPS-----RGKLVGYKIFY 1010  
Qy 1467 TQLRAQCKEWSSECEVQKIQYLYEAHLTET-VYGLKFRWRVQVAASNAGYGNA 1525  
Db 1011 TMTAVEDLEWQTK-----VVGVTESADLINLEKQAQYAVAIAAMVTKGLGL 1058  
Qy 1526 SNWISTKTLPGGPDGPPNVHVVATSPFISISWSBPVITGPTCYLIDVKSVNDDEFNI 1585  
Db 1059 SE-----KATVKMAAPQPMVQPDFYGV-----VNGBEIOV 1088  
Qy 1586 SPKSNNEKNTIEKDLIEFTYSVVITFTGNISAAVYEGKSSAEMIVTLESAPKDP 1645  
Db 1089 ILPQASEE-----YGPISHY-----LIV-----VPEDKA 1113  
Qy 1646 NNMTFOKIPDEVTKFQLTFLPPSQP-----NGNIQVQYALVREDPTAVOI-----HNLS 1696  
Db 1114 N-----LHKIDQ-----FLTEBLLSPRPKLDRLNAPYIAAMFLORNIPTFHLGSETWHP- 1166  
Qy 1697 IIQKNTFTVIAMLEGLGHTYINISYAVNSAGAGPKVPMRITMDIKAPARKTKPTTY 1756  
Db 1167 -----TN-----KLERGKRYIFVRV-----V 1185  
Qy 1757 DATGKLLVSTTITIRMPICYSDRGPINKVQVLATETGAQHDGNTWKYDAYFNKARP 1816  
Db 1186 DTPQKHLVTSPPS-----EFLALDMKEATPGE-----PP 1215  
Qy 1817 YFTNEGFPP-----CTEGKTKFSGNEE-----IYIIGADNACMPCGNEDKICNGPLKPK 1868  
Db 1216 HRPN-----PNVPQTEIIRGK-----NEEPGLMWVG----- 1244  
Qy 1869 QYLKFRATNIMQFTSDSYSDPVKTLGELGSERTVEIILSVTLCLSLILGTAIFAPA 1928  
Db 1245 -----PI-----IAALVLSVCLF-----VYFVR 1263  
Qy 1929 RIRQKQEGTYSQDAEIIDTKLDQLITVADLELKERLTR-----PISKKS 1978  
Db 1264 RRRQPCK-----APDQAAVTRPLMAADLGAAGPAPTDPVDMRINFQTPGMISHPPISIA 1318  
Qy 1979 FLQHVLELCTNNLKFQRESEL-----PKFLQDLSSTADLPWNAKRNRPENKPYNNNR 2035  
Db 1319 LANHVERLKANLKFQSETESTEPGQOFTWDSHNEV-----NPKRNVANTVSY-DHGR 1373  
Qy 2036 VKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMVWETRAKTLVMTQ 2095  
Db 1374 VILPPIEGVGSYINANYCDGVRKHNAYVATQGLQETFGDFWRMCWELKSTIVMTIR 1433

Qy 2096 CPEKGRIRCHOYWPBNDKPVTVFGDIVITIKLMEDVQIDWTIRDLKIERHG--DCMTVRQC 2153  
Db 1434 LEERSRIKCDMWPA--RGTEVYGAMTVTITETQLATVSIITFQIYRNGSNERREIKQL 1491  
Qy 2154 NETAMPGEVGPNSAPLIHFVKLVASRAHDTTPMIVHCSAGVGTGVFIADLHLOHIN 2213  
Db 1492 QFTAWPEDGVDPHPAPFLQFLARLTKSLTPSESGPIIVHCSAGVGTGCVIVIDSLMRMK 1551  
Qy 2214 DHDFVDIYGLVAELSERMCMVQNLQAOYIFLHOCILDLNLSNKGSNQPICFVNVYSALQKWD 2273  
Db 1552 YKTIIDYIGHVTVCLRAQRNMVQTSDDQYIFIHDALLEAVI-CGSTEVPARSLSHNIQKLM 1610  
Qy 2274 SLDAMEGDVELEWE 2287  
Db 1611 QTEPHENITGMEWE 1624

## RESULT 8

Q9EQ17 PRELIMINARY; PRT; 1898 AA.  
AC Q9EQ17; (TREMELrel. 16, Created)  
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)  
DE Tyrosine phosphatase LAR.  
GN Name=Ptpfr;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Thymus;  
RX MEDLINE=21135493; PubMed=11241288;  
DOI=10.1003/1521-4141(200103)31:3<832:AJD-IMM08323.0.CO;2-D;  
RA Tarszowski G., Jankowski A., Hendricks W.A.J., Rolink A.G., Kiselow P.;  
RT "Within the hemopoietic system, LAR phosphatase is a T cell lineage-specific adhesion receptor-like protein whose phosphatase activity appears dispensable for T cell development, repertoire selection and function."  
RT function."  
RL Eur. J. Immunol. 31:832-840 (2001).  
DR EMBL; AF300943; AAG40194.1; --  
DR HSSP; P10586; LLAR.  
DR MGD; MGI:102695; Ptpfr.  
DR GO; GO:0005615; C:extracellular space; TAS.  
DR GO; GO:0016021; C:integral to membrane; TAS.  
DR InterPro; IPR003962; FNIII subd.  
DR InterPro; IPR003961; FN III.  
DR InterPro; IPR008957; FN\_III-like.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003598; IG\_c2.  
DR InterPro; IPR000387; TYR\_phosphatase.  
DR InterPro; IPR000242; Tyr\_PP.  
DR Pfam; PF00041; fn3; 7.  
DR Pfam; PF00102; Y\_phosphatase; 2.  
DR PRINTS; PRO0014; FNTPERIII.  
DR PRINTS; PRO0700; PRTPHPHTASE.  
DR SMART; SM00060; FN3; 8.  
DR SMART; SM00408; IGC2; 2.  
DR SMART; SM00194; PTPC; 2.  
DR PROSITE; PS00853; FN3; 8.  
DR PROSITE; PS00835; IG\_LIKE; 3.  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.  
DR PROSITE; PS00556; TYR\_PHOSPHATASE\_2; 2.  
DR PROSITE; PS00555; TYR\_PHOSPHATASE\_PTP; 2.  
KW Hydrolase.  
SQ SEQUENCE 1898 AA; 211504 MW; BFD48DD11B352A4A CRC64;

Query Match 9.9%; Score 1186; DB 2; Length 1898;  
Best Local Similarity 24.4%; Pred. No. 5.3e-53;  
Matches 477; Conservative 282; Mismatches 713; Indels 484; Gaps 71;

QY 461 PEIWNVEPMVGLYEGSAEMSSDLHSLATFIYNSHDXPNPARRAEDQTSPPVT----- 515  
DB 33 PVFVKVPEDQTLGSEG-----VAFVC-----OATGEPKPRITWKKG 70  
QY 516 --TRNOYITDIAAEOQLSYVIRLVFPF-----TEHMSVSAFTIME--GPPTVLSVRTRQOV 568  
DB 71 KKVSSORFVEIFDDGAGSVLRIOPLRVORDEAIYECTATNSLGEINTSAGLSVEEDQL 130  
QY 569 PS-----SIKIINYKNISSSILLYNDP-----PEYPNGKITHYTYI 605  
DB 131 PSGFPTIDMGPOLKVVKEGRTATMLCAAGGNPDPEISWFKDLPDPVPAASNGRIKQLRSG 190  
QY 606 AMELDTNRAPOITTIIDNSFLITGLKYYTKYKMRVAASHTDGBSSLSSENDIFVRTSEDEP 665  
DB 191 ALQIESSEE-----SDQGYE-----CVATNSAGTRYAPANLYVVRVRVAP 232  
QY 666 ESS--PODEV-----IDVTADEIRLKWSPPEKPNKNGIIIAEVLVKNIDTLYMKN 713  
DB 233 RFSIPSSQVMPGGSVNLTCAVAGAMPYVKK-----MGAELTKEDEMPVGRN 283  
QY 714 TSTTDIILNRLPHTLYNISVRSYTRFGHGNQVSSLSVRSSTVPSA-----PENITYK 769  
DB 284 VLELSNMRS-----ANYTCA-----ISSLGMIEATAQVTVKALPKPPIDLVVT 328  
QY 770 NISSGIEHLSFLPPSPNGIHKYTYLKR--SNGNEERTINTTSLTONIKVLKKTQYI 827  
DB 329 ETTATSVTLTW--DSGNTEPVSYGQYRAAGTDGPFQEVGVASTRYSIGLSPESEYA 386  
QY 828 IEVSASTLKEGVRSAPISLITTEDAPDPPQFSVKQLSGVTVKLSWQPPLEPNGLIY 887  
DB 387 FRVLAVNSIGRPPSBAVARTGEQAPSPPRVQARMLSASTMLVQWPEEPPEFNLVRG 446  
QY 888 YTVY-----VNRSLKTIINTETSL--ELSDLDYNVEYSAYVTASTRFGDGKTG 935  
DB 447 YRYVYTPDSRPLSAMHKN-----TDAGLLTVGSLLPGVYYSRLVLAFTAVGPGPP- 499  
QY 936 SNIISFOTPEGADSPDKVYVYANLSSSS--ILFWTPPSKPNGLIIOYYSVYKNTSGTFM 994  
DB 500 SPTIQVKTOQGVARSAD--FQANAESDTRIQLSMLLPQER-IVKYELVYV-----AAE 551  
QY 995 QNFTLHELNDPDMVTSTIIDKLTIFSYTYFWLTASTSVGNKSSDIIIEVTDQDIPE 1054  
DB 552 DEGOQKHVTFD--PTSSYVLEDLKPDTLRYFQLAARSDLGVG-VFTPTVEAXTAQSTPS 607  
QY 1055 GFVGNLTYESISSTAINVSWPPA--QPNGLV-----FYVSLIQTQPRHVRPLVYTER 1108  
DB 608 APPQKVTCTVSTGTVVRSVWPPPADSRNGIITQYSVAYEAVDGEKRVVDGISEHS 667  
QY 1109 SIYFDNLEKYTDYILKITPSTKEGFSDDTYTAQYIKTEEDVPETSPINTFKNLSSTSVL 1168  
DB 668 SWDLLGLEKWTEYRVWVRAHTDVG-PPPESSPVLRTDEDPGPPRKVEVEPLNSTAVH 726  
QY 1169 LSWDPPV--KPNCAIISYDLTLOGPNENYSFTSDNYIILELSPFTLSFFFAAATKRG 1226  
DB 727 VSWKLPVPNKQHQIRGYQVT-----YVLENGEP----- 756  
QY 1227 LGPSSILFFYTDSDVPLAPPONLTINCTSDFVWLKWSPLPGGIVKVYSPFKIHEHTD 1286  
DB 757 RQQPIL-----QDVMLAEAQ----- 771  
QY 1287 TIYKYNISGFKTBKLVGLEPVSSTYIRVSAPFTKVGNGNQFSNVVFKTTQESVPDVVQNM 1346  
DB 772 -----ETTISGLTPETTYISITVAAYTTKGDGAR-SKPKVVTITGAVPG-RPTM 817  
QY 1347 QCWATSWQSVLKWDPKPKANGIITQMVTVTER-----NSTKVSQDHEMYTPIKLANT 1400  
DB 818 MVSTAMHTALLQWHPKPKELGELLGRLQYRREADEARPNTIDFGDDOQHFVTGLHKA 877  
QY 1401 SYVPKVRASSTAGEBDSCHVSTLPTVPS-VPTNIAESDVQSTSATITWIRPDITL-- 1457  
DB 878 TYVFRLAANKRAGBEFEKBITT-PEDVPSGPPQNLRVTLTSTTELTWDPP--VLAE 934

QY 1458 --GYFONY-----KITTLQRAQCKEWESEBCEVYQKIQYLYEAHLTEETVYGLKKFRWY 1510  
DB 935 RRGXIITNYTVVYRDINSQLELQNVTN-----DTHL-----TLGLKRPDITY 975  
QY 1511 RFQVAASTNAGYGNASNISTKTLPDPDPPEHNVHATSPSISISSEK-----AVI 1565  
DB 976 DIKVRATSKGAGPLSPSISQSRTP-VEQFTKGNFRVAAAMKTSVLLSWEVPSYKSAV- 1033  
QY 1566 TGPTCYLIDVKSVYDNDDEFNISFKISNEENKTIIEKDLIFTRYSVVITAFPTNISAAYVE 1625  
DB 1034 --PKILYNGQSVEVDGHSKML-----IADLPNTIEYSFVL--MNRGSSAGGLQ 1079  
QY 1626 GKSSAEMIVTLESAP-----KDPNNMTFQKI--PDEVTKFQLTFPLPPSPQNGIQV 1676  
DB 1080 HLVSINTAPDLLPQKPLPASAFIEDGRFSLSPQVQDPSLVRWFYVWVVIDRVGNL-- 1137  
QY 1677 YQALVYREDDPTAVOIHNLISIIQKTNFTVIAMLEGKLGHTYNISVAVNSAGACPKVPM 1736  
DB 1138 --LAPRWTPPELEDEL-----LEAIEQSEBKQ----- 1164  
QY 1737 RITMDIKAPARPKTPPIYDATGKLLVTSTTITIRMPICYDDDDHGPINKVQVLATETG 1796  
DB 1165 -----RRRRROAERLKPVYAAQVDVLPDTFTLGDK----- 1194  
QY 1797 AQHGDGNTWYDAYFNKARPYPFTNEGFPNPPCTEGTKFSGNEEYIIGADNACMIPGNE 1856  
DB 1195 -----KSYRGFYNR----- 1203  
QY 1857 DKICNGPLPKKQY-----LPKFRATNIMGQFSDSDYDPVKTLGEGLSERTVEIILSVT- 1911  
DB 1204 -----PLSPDLISYQCFVLASKKEPMDQRYASSPYSDEIVVQVTPAQOQSEBEMLVWTG 1257  
QY 1912 --LCILSIITLGTALPAPARIRQKQEGGTYPQDAEIIIDTKLKLQDLITVADLELKDER 1969  
DB 1258 PVLAVILIIILVIAILLFKKR-----THSPSSKDEQSIGLK-DSLLAHSDDPVEMRR 1309  
QY 1970 LT-----RPISKKSFLQHVHELCTNNLKFQEFSEL-----PKFLODLSSTDADLPW 2017  
DB 1310 LNYQTPGMRDHPPIPTIDLADNIERLKANDGLKFSQEYESIDPGQOFTWENSEV---- 1365  
QY 2018 NPAKRRFPNIKPVNNNRVKLIADASVPGSDYINASYISGLCPNFIATQGPLGTVGD 2077  
DB 1366 NPKPKRIADVIAY-DHSRVLLTSIDGVPGSDYINANYIDGKQKQNIATQGPLPETMGD 1424  
QY 2078 FWRVWETRAKTLVMLTQCFEKGIRCHOYWPEDNKNPVTVPFGDIVITIKLMEVQI-DWTI 2136  
DB 1425 FWRVWEOQTATVWMTRELSRVKCDQYWPV--RGTYGLIQT-LVDTVELATYTM 1481  
QY 2137 RDLKIERHG--DCMTVRCQNFAMPEHGVPENSAPLIHFVKLVRASRAHDTTPMIVHCSA 2194  
DB 1482 RTFALHKSGSSEKRELROQPMAPDHPGVPEYPTPIFLAFLRRVKACNPLDAGPMVHCSA 1541  
QY 2195 GVGRTGVFTALDHLQIHNDHDFVDIYGLVAELSERMCMQVONLAQYIFLHQCILDLSN 2254  
DB 1542 GVGRTGCFIVIDAMLERMKEHTVDIYGHVTCMSORNTWVQTEDQYVFIHEALLR-AAM 1600  
QY 2255 KGSNOPICFVNTYSALQKM-----DSLDAEGDVEL 2284  
DB 1601 CGHTEVLARNLTAHQKLGQVPPGESVTAMELEFKL 1636  
RESULT 9  
Q64604  
ID Q64604 PRELIMINARY; PRT; 1898 AA.  
AC Q64604; Q63294; Q63295; Q63296;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Protein-tyrosine phosphatase, receptor-type, F polypeptide precursor  
DE (EC 3.1.3.48) (LAR protein) (Leukocyte antigen related) (Leukocyte  
DE common antigen-related phosphatase) (Protein-tyrosine-phosphatase)  
DE (Phosphotyrosine phosphatase) (PTPASE).  
GN Name=Lar;

OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;  
RX MEDLINE=94347119; PubMed=8068021;  
RA Zhang W.R., Hashimoto N., Ahmad F., Ding W., Goldstein B.J.;  
RT "Molecular cloning and expression of a unique receptor-like protein-  
tyrosine-phosphatase in the leucocyte-common-antigen-related phosphate  
family.";  
RL Biochem. J. 302:39-47(1994).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS PRLAR4.0 AND PRLAR631).  
RC STRAIN=SPAGUE-DAWLEY; TISSUE=HIPPOCAMPUS;  
RX MEDLINE=95146548; PubMed=7844155; DOI=10.1083/jcb.128.3.415;  
RA Zhang J.S., Longo F.M.;  
RT "LAR tyrosine phosphatase receptor: alternative splicing is  
preferential to the nervous system, coordinated with cell growth and  
generates novel isoforms containing extensive CAG repeats.";  
RL J. Cell Biol. 128:415-431(1995).  
RN [3]  
RP SEQUENCE OF 1035-1898 FROM N.A., AND MUTAGENESIS.  
RC TISSUE=HYPOTHALAMUS;  
RX MEDLINE=92011772; PubMed=1918076;  
RA Pot D.A., Woodford T.A., Remboutsika E., Haun R.S., Dixon J.E.;  
RT "Cloning, bacterial expression, purification, and characterization of  
the cytoplasmic domain of rat LAR, a receptor-like protein tyrosine  
phosphatase.";  
RL J. Biol. Chem. 266:19688-19696(1991).  
CC -|- FUNCTION: IT IS POSSIBLE THAT LAR IS A CELL ADHESION RECEPTOR. IT  
POSSESSES AN INTRINSIC PROTEIN TYROSINE PHOSPHATASE ACTIVITY  
(PTPASE). IT MODULATES SIGNALING BY THE INSULIN, EPIDERMAL GROWTH  
FACTOR, AND HEPATOCYTE GROWTH FACTOR TYROSINE KINASE RECEPTORS.  
MAY PLAY A ROLE IN NEURITE OUTGROWTH AND/OR CELL-CELL INTERACTIONS  
MODULATING SYNAPSE FORMATION. MAY INFLUENCE CELL MOTILITY.  
CC -|- FUNCTION: THE FIRST PTPASE DOMAIN HAS ENZYMACTIC ACTIVITY, WHILE  
THE SECOND ONE SEEMS TO AFFECT THE SUBSTRATE SPECIFICITY OF THE  
THE FIRST ONE (BY SIMILARITY).  
CC -|- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O = PROTEIN  
TYROSINE + PHOSPHATE.  
CC -|- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
CC -|- ALTERNATIVE PRODUCTS:  
CC Event=alternative splicing; Named isoforms=5;  
CC Comment=Additional isoforms seem to exist;  
CC Name=PRLAR;  
CC IsoId=Q64604-1; Sequence=Displayed;  
CC Name=PRLAR631;  
CC IsoId=Q64604-2; Sequence=VSP\_050409, VSP\_050410, VSP\_050413,  
VSP\_050415;  
CC Name=PRLAR4.0;  
CC IsoId=Q64604-3; Sequence=VSP\_050415;  
CC Name=PRLAR18.1;  
CC IsoId=Q64604-4; Sequence=VSP\_050414, VSP\_050415;  
CC Name=PRLARCB9;  
CC IsoId=Q64604-5; Sequence=VSP\_050411, VSP\_050412;  
CC -|- TISSUE SPECIFICITY: EXPRESSED PREFERENTIALLY IN NEURAL OR  
NEUROMUSCULAR TISSUE.  
CC -|- DEVELOPMENTAL STAGE: THE ALTERNATIVELY SPLICED ISOFORMS ARE  
DEVELOPMENTALLY REGULATED.  
CC -|- SIMILARITY: EXTRACELLULAR REGION TYPIC OF A CAM FAMILY (3 IG-LIKE  
DOMAINS FOLLOWED BY 8 FIBRONECTIN TYPE III-LIKE DOMAINS), AND A  
CYTOPLASMIC REGION COMPOSED OF TWO PTPASE DOMAINS (BY SIMILARITY).  
DR EMBL; L11586; AAC37655.1; -;  
DR EMBL; M60103; AAA41510.1; -;  
DR EMBL; X83505; CAA58495.1; -;  
DR EMBL; X83546; CAA58537.1; -;  
DR FIR; S46216; S46216.  
DR HSSP; P10586; ILAR.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR GO; GO:0005515; F:protein binding; IEA.

DR GO; GO:0004725; F:protein tyrosine phosphatase activity; IEA.  
DR GO; GO:0004872; F:receptor activity; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.  
DR InterPro; IPR003962; FNIII\_subd.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR008957; FN\_III-like.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003598; IG\_c2.  
DR InterPro; IPR001005; MY5\_DNA\_binding.  
DR InterPro; IPR000387; TYR\_phosphatase.  
DR InterPro; IPR000242; Tyr\_PP.  
DR Pfam; PF00041; fn3; 7.  
DR Pfam; PF00102; Y\_phosphatase; 2.  
DR PRINTS; PR00014; ENTYPETII.  
DR PRINTS; PR00700; PRTPHPHTASE.  
DR SMART; SM00060; FN3; 8.  
DR SMART; SM00408; IGc2; 2.  
DR SMART; SM00194; PTPC; 2.  
DR PROSITE; PS50853; FN3; 8.  
DR PROSITE; PS50835; IG\_LIKE; 3.  
DR PROSITE; PS00037; MYE\_1; UNKNOWN 1.  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.  
DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 2.  
DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 2.  
KW Alternative splicing; Cell adhesion; Glycoprotein; Hydrolase;  
KW Immunoglobulin domain; Receptor; Repeat; Signal; Transmembrane.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 1898 PROTEIN-TYROSINE PHOSPHATASE, RECEPTOR-  
TYPE, F POLYPEPTIDE.  
FT DOMAIN 28 1254 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 1165 1169 POLY-ARG.  
FT TRANSMEM 1255 1275 POTENTIAL.  
FT DOMAIN 1276 1898 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 47 114 IG-LIKE C2-TYPE DOMAIN.  
FT DOMAIN 149 214 IG-LIKE C2-TYPE DOMAIN.  
FT DOMAIN 246 305 IG-LIKE C2-TYPE DOMAIN.  
FT DOMAIN 317 ? FIBRONECTIN TYPE-III.  
FT DOMAIN 435 ? FIBRONECTIN TYPE-III.  
FT DOMAIN 1361 1607 PROTEIN-TYROSINE PHOSPHATASE.  
FT DOMAIN 1650 1898 PROTEIN-TYROSINE PHOSPHATASE.  
FT ACT\_SITE 1539 1539 BY SIMILARITY.  
FT ACT\_SITE 1830 1830 BY SIMILARITY.  
FT DISULFID 54 107 POTENTIAL.  
FT DISULFID 156 207 POTENTIAL.  
FT CARBOHYD 117 117 POTENTIAL.  
FT CARBOHYD 253 298 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 295 295 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 721 721 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 941 941 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 957 957 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT VARSPLIC 600 707 CTAQSTPSAPPOKVCSTGSTRVSVWVPPADSRNGIIT  
QVSVAEYAVDGRKRVVVDGSRHSSMDLGLKWKTEYR  
VWVRATDVGPSPSPVLVRTDEDV -> RTAQM (in  
isoform PRLAR631).  
FT VARSPLIC 771 771 FTId=VSP\_050409.  
Q -> QMRPESEDI (in isoform PRLAR631).  
FT VARSPLIC 810 810 /FTId=VSP\_050410.  
A -> AGE (in isoform PRLARCB9).  
FT VARSPLIC 811 1898 /FTId=VSP\_050411.  
Missing (in isoform PRLARCB9).  
FT VARSPLIC 812 1004 /FTId=VSP\_050412.  
Missing (in isoform PRLAR631).  
FT VARSPLIC 1276 1276 /FTId=VSP\_050413.  
K -> KSKQE (in isoform PRLAR18.1).  
FT VARSPLIC 1316 1316 /FTId=VSP\_050414.  
G -> GSAFSPCNIS (in isoform PRLAR4.0.  
isoform PRLAR631 and isoform PRLAR18.1).  
FT MUTAGEN 1539 C->S; LOSS OF ACTIVITY.  
FT CONFLICT 777 777 G -> S (IN REF. 2).  
FT CONFLICT 1073 1073 T -> S (IN REF. 2 AND 3).

FT CONFLICT 1434 1434 I -> T (IN REF. 2 AND 3).  
FT CONFLICT 1639 1639 G -> N (IN REF. 2 AND 3).  
FT CONFLICT 1643 1644 RA -> HT (IN REF. 2 AND 3).  
SQ SEQUENCE 1898 AA; 211492 MW; DF4D7E46F5896F4B CRC64;

Query Match  
Best Local Similarity 9.9%; Score 1181; DB 2; Length 1898;  
Matches 478; Conservative 273; Mismatches 704; Indels 518; Gaps 72;

QY 461 PEIVNIVBVMGLYGEAEMSSDLHLATFIYNHDPDKNPPARNRAEDQTSPPVT-----515  
DB 33 PVPVKVPEQIGLGG-----VAFVC-----QATGPKPRIITWKKG 70

QY 516 --TRNOYITDIAEQLSVIRLVPF-----THEMISVSAPTINGE--GPPTVLSVRTRQQV 568  
DB 71 KKVSSORFVIEPDDGAGSVLRIOPLRVORDEAIECTATNSIGINTSAGLSVLEEDQL 130

QY 569 PS-----STKIINYKNISSILLYWDP-----PEYPNGKITHYTIY 605  
DB 131 PSGFPTIDMGPOQKVVEKARTATMLCAAGNPDPEISWFKDPLVPDPASSNGRIKOLRS 190

QY 606 AMELDNRAFOIITIDNSFLITGLKYYTKYKMRVAASHTDGSLSSEENDIFVRTSEDEP 665  
DB 191 ALQIESSE-----SDQKYE-----CVATNSAGTRYAPANLYVVRVAP 232

QY 666 EGS--PODVEVIDTADIELKWSPEKPGNGIIIAVEVLYKNIDTLYMKNSTSTDIILRN 723  
DB 233 RFSIPSSQEVN-----PGG-----NVNLTCAVAGA 258

QY 724 LRPHLYNLSVRSYTRFGH--GNQVSSLLSVRTSETVPDSPAENITYKNISS-----773  
DB 259 PMPYVMMGABELTKEDEMPVGRNVLSESNMRS-----ANYTCVAISSLGMEAT 310

QY 774 GEIELSFLPSSPNGIHKYT--IYLRKSNNGEE-----RTINT-----T 811  
DB 311 AQTVKALPKPPIIDLVTETATSVLTWDSNGTEPVSFYQYRAAGTDGPFQFVDGVA 370

QY 812 SLTONIKVLKYYQYIIEVSASLTKGEVRSAPISILTEEDAPDPPQDFSVKQLSGVTV 871  
DB 371 STRYSIGGLSPFSEYAFRLAVNSIGKGPSEAVRARTGEQAPSSPRRVRQARMLSASTM 430

QY 872 KLSWQPLEPBGHILYTYV-----VNRSSLKINTVETSL--ELSDLDYNYEY 919  
DB 431 LVQEBEPPEPGLVGRVYVYTPDSRRPLSAMHKN-----TDAGLLTVGSLLPGITY 484

QY 920 SAVVTASTRGDCGKTSNIIISFOTPEGAPSDRPKOVVYANLSSSSIIILFWTPSKENGII 979  
DB 485 SURVLAFTAVGDGP--SPTIQVKTQGVPAQ--PADFOAKAESDTRIQLSWLLPPQBR--II 541

QY 980 QYVSYYRNTSGTFMQNFIHLTNDPDMNTVSTIIDKLITFSYTYFTWLTASTSVGNNGK 1039  
DB 542 KYELVW-----AAEDBGGQHKVTPD--PTSSYTLLEDKPDLYHFQLAARSDLGVG-V 592

QY 1040 SSDIIEVYTDQDIPGFGVNLTVESISSAINVSWVPPA--QBNGLV-----FYVSLIILQ 1093  
DB 593 FPTVEACTAQSTPSAPPQKVTCSGTSITVRVSWVPPADSRNGIITQVSAVEADGE 652

QY 1094 QTPHRVRPLVTVERSIYFNDLEKYTDYILKITPSTEGFSDYTAQLVIXTEEDVPETS 1153  
DB 653 DRKRHRVVDGISREHSSWDLGLEKTEYRVMVRAHTDVG--PGPESPVLVTRDDEVPSPG 711

QY 1154 PIINTFKNLSSTSVLSWDDPPV--KPNGATISYDLTLQGNENYSPIITSNYIILELSP 1211  
DB 712 PRKVEPEPLNSTAVHVSXKLPVKNQHQGIRGQVT-----VYRLENGEP 756

QY 1212 FTLVSPFAARTKGLGPSILFFYTDSEVPLAPPQNLTLINCTSDFWLKNWSPSPGPG 1271  
DB 757 -----RGQPII-----QVWMLAEAQ-----771

QY 1272 IVKVSFKIHEHTDIYKYNISGFKTEAKVLEPVSITYSIRVSAFTKVGNQNPVNV 1331  
DB 772 -----ETTISGLTPTTYSITVAAYTTKGDGAR--SKPK 803

RESULT 10

Q6PG86

ID Q6PG86

AC Q6PG86;

PRELIMINARY;

PRT; 1529 AA.

QY 1332 KETTOESVDPVVONMOCMATSWQSVLVKWDPPKANGIITQYMTVVER-----NSTKVS 1385  
DB 804 VVTTTGAVEG-RPTMMVSTAMHTALLQHPHKLPELGLGYRLQVRRDAEPNTPDFG 862

QY 1386 PODHMYTTFIKLANTSYPVKVRASTAGSGDESBTCHVSTLPETVPS-VPTNIAFSDVQST 1444  
DB 863 KDDQHTVTGLHKGATYIFRLAAKNRAGPGBEFKEIIT-PEDAPSGPQNLRAVTGLTTS 921

QY 1445 SATLTWIRPDTIL---GYFQNYKITQLRAOKCKEWESEECVEYQIKIYLV- 1493  
DB 922 TTLEAWDP--VLAERNGRITNTYTV-----YRDISQHELQNVTVGD 961

QY 1494 AHLTETVYGLKPFMYREFOVAASNAGYGNASNMISTKTLPCPPDPENVHVATSPF 1553  
DB 962 VHL---TLGLKPDPTTYDIKVRHRTSKGAGPLSPSQSRTP-MEQVAFKFRVAAMKT 1017

QY 1554 STISWSE----AVITGTCYLDVKSVDNDEFNIFISPKNEENKTIEIKOLEIFTRY 1608  
DB 1018 SVLLSWEVDSYKSAV---PFKILYNGQSVEDVGHSMRKL-----IADLPQNTY 1064

QY 1609 SVVITAFGTNISAAYVEGKSSAEMIVTTLESAP-----KOPPNMTFOKI--PDEVTK 1659  
DB 1065 SPVL--MNRGTSAGGLQHLVSIRTPADLLPQKPLPASAFIEDGRFSLSPQVQDPSLVR 1122

QY 1660 FQITFLPPSPQNGNTQVYQALVYREDDPTAVQIHLNLSIIQNTFTVIAMLEGLKGHTYN 1719  
DB 1123 FYIVVVPIDRVGNL-----LAPRWSTPBELEDEL-----LEAIEGEEKQ 1164

QY 1720 ISVAVNSAGAPKVPMTITMDIKAPARPKTPTIYDATKLLVTSTTITRMPICYYS 1779  
DB 1165 -----RRRRQARLKPVAQVDELPTFTL----- 1191

QY 1780 DDHGPIKNQVQLATETGAQHDGNVTWKYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNE 1839  
DB 1192 ---GDCKN-----YRGPYNR----- 1203

QY 1840 EYIIGADNACMIPGNEDKICNGPLKPKQY-----LFKFRATNIMQFTDSYSDPVKTL 1895  
DB 1204 -----PLSPDLSYQCFVLASIKBPMQDKRQYASSPSYSDIVVQ 1240

QY 1896 GEGLSERTVEIILSVT---LCILSIILGTAFAPARIROKQEGGYSPQDAEIIIDTKL 1952  
DB 1241 VTPAQOQEBEPMLWVTGPFVLAVILIIILVIALLLFKRK-----THSPSSKDEQSIGL 1293

QY 1953 KLDQILITVADLELXDERLT-----RPISSKPSLQHVLELCTNNNLKFOEFSEL-- 2001  
DB 1294 K-DSLLAHSSDPVEMRLNYQTPGMRDHPPIITDLADNIERLKANDGLKFSQEYSDIP 1352

QY 2002 -PKFLODLSSTADADLPWNRKARFPNFKPNNNNRVKLJADASVPGSDVINASYLC 2060  
DB 1353 GQOFTWENSSEV---NKPKNRYANVIAY-DHSRVLLTSIDGVPGSVDINANYIDGYRK 1407

QY 2061 PNEFIATQGLPCTGDFWRWVWETRAKTLVMTQCFEKGRIRCHOYWPEDNKPVTVFGD 2120  
DB 1408 QNAYIATQGLPETGDFWRWVWVORATVVMTRLEESRVKCDQYWFA--RGTYTYGL 1465

QY 2121 IVITIKLMEVDQI--DWTIRDLKIERHG--DCMTVROCNFTAWPEHGVHPENSAPLIHFVKLV 2177  
DB 1466 IQVT-LVDVTELATYMTFTALHKSSEKRELQOFQWAPDHGVPEYPTPIALFRRV 1524

QY 2178 RASRAHDTTPMIVHCSAGVGRGTGVTIADLHQLQHINDHFDVDIYGVLAELSRMCMQVN 2237  
DB 1525 KACNPLDAGPVMVHCSAGVGRGTGCFIVIDAMLERMKHEKTVDIYGHVTCWRSQRNTVQT 1584

QY 2238 LAQYIFLHQICILDLNKSQSNQPICFVNTSALQKM-----DSLDAEGDVEL 2284  
DB 1585 EDQYVPIHEALLE-AAMCGHTEVLARNLYAHTIQLGQVPPGSSVTAMELEFKL 1636













```
Db 1056 SYFPPV-----NATQKIDGLVGNHYIFRIQAKSALGVGAEREHQTPIIAPPVPS 1110
QY 1752 PTPIDYATGKLLVTSTTIIRMPICYSDDHGHPIKQV-VLAPETGAQHDG-NVTKYWD- 1808
Db 1111 VTFL-----EVSRTSTBIISPRQGYFSNAGHMVRSYITIIAARDVGNASGLEMPQSDV 1165
QY 1809 -AY-----FNKARPYPNNGFPNPPCT-BOKTKFSNGEIEYIIGADNACWIPGNEDK--- 1858
Db 1166 QAYTWLWLPYQAIEPY-----NPFITSGSRKSLSEAEHFTGTATNC-----DKHOA 1211
QY 1859 -ICNGPLPKPKQYLFKFRATNMGQFTSDSDYSDPVVTIGLSEGLSERTVEIILSVTLCLIS 1917
Db 1212 GYCNGLPERAGTTRIKIRAFTEDEKFTDVIYSSPIT-----ERSDTVIVAAT---VSA 1262
QY 1918 ILGTAIFAFAPIRQKQEGGTYSPQDAEIIDTKLX-----DGLIIVADLELKDRLTRP 1973
Db 1263 VLIAMVVLV-----VYQHRCLQRIEASKLARMQDELAALPGSYITP---NRP 1308
QY 1974 ISKSKFLOHVEHLCNTNNLKFQEFSELPKFLQDLSSDADLPWNRKAKRPNPNKPYNN 2033
Db 1309 VHKDFSEHYRIMSADSDFRSEFEELKHVGRDQACSFANLPNRPKNRFTNLPY-DH 1367
QY 2034 NRKVLITADSVRGSVDYINASYISGLYCPNEFIATQGLPCTVCGDFWRVWETRAKLVML 2093
Db 1368 SRFLQPVDDGSDYINANYMPGHSNPREFIVTQGLHSTREBFWRMCWESNRAIVML 1427
QY 2094 TOCFEGRIRCHQWYEDNKPVTVFGDIIVTKLMEDVQIDWTIRDLKIRHGDGCMTVROC 2153
Db 1428 TRCFEGRKCDQYVPD-RVAMFYGDIKVQLIIDTHYDWSISEFWVSNCESRIMRHF 1486
QY 2154 NFTAWEHGVENSAPLHIFVKLVASRAHDTTPMTVHCSAGVGRGVFIALDHLTQHJN 2213
Db 1487 HFTWPDGFPVPEPQSLVFRVAFRDVIGTDMRPIIVHCSAGVGRSGTFFIALDRILQIH 1546
QY 2214 DHDFVDIYGLVAELSERMCWQNLQAYIFLHCILDLISNK 2255
Db 1547 KSDYVDIFGIVFAMRKERVFMVQTEQQYVCIIHQCLLAVLEGK 1588

RESULT 14
Q9W4F5
ID Q9W4F5 PRELIMINARY; PRT; 1767 AA.
AC Q9W4F5;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE CG6899-PA.
GN Name=Ptp4E; ORFNames=CG6899;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=107311132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J.J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
```

```
RA Clodek A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Kallal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupeki M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).
RN [3]
SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).
RN [4]
SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tuzy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).
RN [5]
SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL, AE003432; AAF45998.1; --
DR FIC, A49502; A49502.
DR FIC, B49502; B49502.
DR HSP, P10586; ILAR.
DR IntAct; Q9W4F5; --
DR FlyBase; FBgn004368; Ptp4E.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004725; F:protein tyrosine phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
```







|    |      |  |      |
|----|------|--|------|
| Qy | 568  | -VPSSIKIINIYKNI-----SSSSILLYWDPPEYPNGKITHYTIYAMELDTNRA         | 614  |
| Db | 304  | GVIEAIAQITVKALPKPGTPVVTTESTATSITLTWDSGN--PGPVSYIIHQHPKNSEEP    | 361  |
| Qy | 615  | FOITPID----NSFLITGLKKYTKYKVRVAASHTDGSSLSSENDLIVRTSEDEPSSPQ     | 670  |
| Db | 362  | YK--EIDGIAITRYSVAGLSFYSDYEFVRVAVNIGRPASB--PVLTOTSEQTSPSAPR     | 417  |
| Qy | 671  | DVEVIDVTAOBIRLKWSPPEKPGIIITAYEVLVKNIDITLYMKNTSTTDI-----ILRLN   | 724  |
| Db | 418  | DVQARMLSSUTILVQWKEPEPNEGIOGYRYVYTMPTQHVNNWKNHNVADSOITIGNL      | 477  |
| Qy | 725  | RPHTLYNIVRSYTRFCHGNQVSGLLSVRTSETPVDSAPENITYKNISSGEIBLSFLPPS    | 784  |
| Db | 478  | VPQKTSYVKVLAFTSIGDG-PLSSDIQVITQGV-CPQLNFKFAEPESSETILLSWTPPR    | 535  |
| Qy | 785  | SPNGIILKKYIYLKRSN-GNEERTINTSITQNIKVLKKYQYIIIEVSASTLKEGVRSA     | 843  |
| Db | 536  | SDT--IASYELVDVDRDGOGEORITIEPGETSYRLQGLKPNLSLYFRLSATSPQGLGASTA  | 593  |
| Qy | 844  | PISILTEEDAPDSPQDFSKVQLSGVTVKLSWOPPLEPNEGIIILYYTVVWNRSLKTINV    | 903  |
| Db | 594  | EISARTWQ-KPSAPPQDISCTSPSTSYILVSWQPP-----627                    |      |
| Qy | 904  | TETSLELSDLDYNVVEYSAYVASTRFGDGKTKGNIISFQPEGAPSPPKDVVYANLSSS     | 963  |
| Db | 628  | -----627   |      |
| Qy | 964  | SIILFWTPSPKNGIIQVYSVYVYRNTSGTFMQNFTLHELHNDPDMTVSTIIDLKTI PSY   | 1023 |
| Db | 628  | -----PVEKQNGIITYESLUKAAVDG----EDYKPEHIIGNSSD-TTKYLLLEQLEKWT E  | 676  |
| Qy | 1024 | YTFWLTASTSVGNKNSDIIIEVYTODIPEGFVGNLYTESISSTAINVSW---VPPAQP     | 1080 |
| Db | 677  | YRITVTAHTDVGPWPELSVL-IRTEDEVSPGPRKVEAEVAVNATAKVWSRSPVENKQH     | 735  |
| Qy | 1081 | NLGVFYVYVSLIQT--PRH--VRPPLVTVERSIYFDNLKYDYILKITPSTTEKGFSDT     | 1136 |
| Db | 736  | GQIRGYQVHYVKMENGPKSAMLKDVMLADAQMIISGLQPETSYSLTVAYTATTKGDGAR    | 795  |
| Qy | 1137 | YTAQLYIKTEEDVPETSPINTFNKLSSTSVLLSWDPPKNGAIIISYDLTLQGPENYS      | 1196 |
| Db | 796  | SKPKL-VSTTGSVPCKPRLV-INHTQMNTALIQWHPVDVTFG-----LQYRLKFG        | 844  |
| Qy | 1197 | FITSDNYIIIEELSGSPFTLYSFFAAARTRKGLGSPSSILFFYTDSEVSLAPPQNTLINCTS | 1256 |
| Db | 845  | -----RKDEPLTTLTFEKE-----860                                    |      |
| Qy | 1257 | DFWVLKWSPLPGIIVKVSFKIHEHETDIITYKNIISGFKTEAKLVGLEPVSYSIRVS      | 1316 |
| Db | 861  | -----DHFTATDIHKGAS-----YVFRLS                                  | 879  |
| Qy | 1317 | AFTKVGNGNOFSNVKFTTOESVP-DVVQNMOCMATSWQSVLVKWDPPKKA--NGIITQY    | 1373 |
| Db | 880  | ARNKVGFGEEM--VKEISVPEEIIPTGPQNLHSEGTSTSVQLSWQPPVLAERNGVITKY    | 937  |
| Qy | 1374 | MVTVERNSTKVSPODHM-----YTFIKLLANTSYVFKVRASTSAGEDSES-TCHVSTL     | 1425 |
| Db | 938  | TLLYRDINVPLLPMEHLIVPADTSMTLTGLKSDTTYDVKVRHAHTSKRRPGAYSPOFRTL   | 997  |
| Qy | 1426 | PETVPSVPTNIAFSDVOSTSATITWIRPDIILGFQNY-----KITTLRAQCKEWES       | 1479 |
| Db | 998  | PVDQSMFAKPHVKAVMKTISVLLSWEIP-----ENYNPAILSKFFMMMDGKQWEE---     | 1047 |
| Qy | 1480 | EECVEYOKIQVLYEAHLTEETVYGLKFRWYRFOVAASTNAG--YGNASNWISTKITLPGP   | 1537 |
| Db | 1048 | -----VDGRATQKILVNLKPEKSYSP---ALTNRGNSAGLQHRVTAKTAP--           | 1090 |
| Qy | 1538 | PDGPPENVHVATSPFSSISISWSBPAVITGPTCVLIDVKSVDNDEFNISI-FKSNENK-    | 1595 |
| Db | 1091 | -----DVLRTKP-----AFIGKTNLDGMI                                  | 1109 |

|    |      |  |   |      |
|----|------|--|---|------|
| QY | 1596 | TIEIKOL---   | EIFTRYSVWITAFGNTISAAAYVEGKSSAEIMVTTLSEAPKDPPNNMTTFQK    | 1652 |
| Db | 1110 | TVQLPDPVANENIKGYIIIVPLK-   | -----KS R G K P I - - - - - K P W E S D P E M D E L D E | 1153 |
| QY | 1653 | IPDEVTKFOLTFLPPSPQNGNI   | QVYQALVYREDDPTAVQHNLHSIIQKTNTFTVIAMLEGL                 | 1712 |
| Db | 1154 | LKKEISRKR-   | -----PYIAAHFDVLPTFTL--                                  | 1162 |
| QY | 1713 | KGGHTYNISVVYVNSAGSAGPKVPMRI                                      | TWDIKAPARPKTPTPIYDATGKLLVSTTTIIR                        | 1772 |
| Db | 1163 | -----RSIRYGREVELK-   | -----PYIAAHFDVLPTFTL--                                  | 1190 |
| QY | 1773 | MPICVYSDDHGPINKNVQLATETGAQHDGNVTWKYDAYFNKARPYFTNNEGFPNPCTEGK     | 1832  |      |
| Db | 1191 | -----GDD-  | -----KH Y G - - - - - G F T N - - - - - K               | 1202 |
| QY | 1833 | TKFGNEEIIYIGA-   | -DNACMI PGNEDEKI CNGPLKP KKQYL FKFRATNIMGQTDSDYS        | 1890 |
| Db | 1203 | QLOSGOEYFFVFLVAMDHA-   | -----ESKM-----VATS PYSD                                 | 1233 |
| QY | 1891 | PVKTLG----   | -----BGLSBRTVEIILSVTLCILSIILLGTATFAFARIQK-----          | 1933 |
| Db | 1234 | PVSMDLDPQPIIDEEGLI----   | -----WVGVPVLA VVFIIICIVIALLYLKRAESDRKSS                 | 1288 |
| QY | 1934 | ---QKEGGTYSPQDAEIIDTKLKUDQLITVADLEKLDERLTRPISKSKSFLOHVEELCTNN    | 1990  |      |
| Db | 1289 | LPNSKEVPFSHPHTD-----   | PVELRL-----NFOTPGMASHPPPILELDADHIERLKAND                | 1338 |
| QY | 1991 | NLKQEPEFSEL PKFLODLSSTDADLPNWRAKRPFNIKPYNNNRVKLIADASVPGSDYI      | 2050  |      |
| Db | 1339 | NLKYSQBYESIDPG-QOFTWEHSNLEVNKKPNRYANVIAY-DHSRVLLSAIEGLPGSDYV     | 1396  |      |
| QY | 2051 | NASTIGYLCNBEFIATQGPLPGVGFWMVMETRAKTIVMLTCQCEKGRIRCHQWPE          | 2110  |      |
| Db | 1397 | NANYIDGVYRKONAIATQGS LPETFGDFWMIWEQEA-TVVMVMTKLEERSVKCDQWP-      | 1454  |      |
| QY | 2111 | DNKPVTVPFGDIVITKLMEDVOIDMTWTRIDLKIERHG--DCMTVROCNTAMPHEGVBPNSA   | 2168  |      |
| Db | 1455 | -SRGTETHGLVQVT-LLDTVHELHLCPDICTLNNGSSEKRRKVRFQFTAMPDPHGVPBHPT    | 1512  |      |
| QY | 2169 | PLIHFKVLVRASRAH--DTTPMIVHC SAGVRCGTGFVALDHLTOHNDHPVDVIYGLVAE     | 2226  |      |
| Db | 1513 | PVPSF--LTESQNLPHPDAGPMVHC SAGVRCGTGCFIVIDAMLERI KH EKTVDIYIGHVTL | 1570  |      |
| QY | 2227 | LRSRMCVMQNLAQYIFILHOCILDLLSNKGNSNQPICFVNYSALQMDSLAMEGDVLEW       | 2286  |      |
| Db | 1571 | MEAQRNYMVQTEDQYIFIHDALEAVTCGTEVPARNL-YAYIQKLQTIETGENVTGMEL       | 1629  |      |
| QY | 2287 | E  | 2287  |      |
| Db | 1630 | E  | 1630  |      |

Search completed: June 30, 2005, 15:04:22  
Job time : 308 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 15:04:38 ; Search time 241 Seconds  
(without alignments)  
3667.457 Million cell updates/sec

Title: US-10-673-885-2  
Perfect score: 11985  
Sequence: 1 MDPLIFLLFTGTSETQVD.....MDSLAMEGDVLEWEETM 2291

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1722976 seqs, 385795295 residues

Total number of hits satisfying chosen parameters: 1722976

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pdb.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pdb.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pdb.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pdb.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pdb.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pdb.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pdb.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pdb.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pdb.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pdb.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pdb.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pdb.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pdb.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pdb.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pdb.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pdb.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10E\_PUBCOMB.pdb.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US10F\_PUBCOMB.pdb.\*  
19: /cgn2\_6/ptodata/2/pubpaa/US10G\_PUBCOMB.pdb.\*  
20: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pdb.\*  
21: /cgn2\_6/ptodata/2/pubpaa/US11\_PUBCOMB.pdb.\*  
22: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pdb.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score   | Query Match | Length | ID                  | Description       |
|------------|---------|-------------|--------|---------------------|-------------------|
| 1          | 11985   | 100.0       | 2291   | 10 US-09-822-871-2  | Sequence 2, Appli |
| 2          | 11985   | 100.0       | 2291   | 15 US-10-673-885-2  | Sequence 2, Appli |
| 3          | 11807.5 | 98.5        | 2300   | 15 US-10-087-684-10 | Sequence 10, Appl |
| 4          | 11807.5 | 98.5        | 2300   | 15 US-10-218-779-10 | Sequence 10, Appl |
| 5          | 11797   | 98.4        | 2299   | 15 US-10-466-759-2  | Sequence 2, Appli |
| 6          | 11191   | 93.4        | 2281   | 15 US-10-087-684-6  | Sequence 6, Appli |
| 7          | 11191   | 93.4        | 2281   | 15 US-10-218-779-6  | Sequence 6, Appli |
| 8          | 10536   | 87.9        | 2301   | 10 US-09-822-871-4  | Sequence 4, Appli |
| 9          | 10536   | 87.9        | 2301   | 15 US-10-673-885-4  | Sequence 4, Appli |
| 10         | 10536   | 87.9        | 2302   | 15 US-10-087-684-37 | Sequence 37, Appl |
| 11         | 10536   | 87.9        | 2302   | 15 US-10-218-779-37 | Sequence 37, Appl |

|    |        |      |      |    |                     |                    |
|----|--------|------|------|----|---------------------|--------------------|
| 12 | 4535.5 | 37.8 | 898  | 14 | US-10-314-232-22    | Sequence 22, Appli |
| 13 | 4488   | 37.4 | 855  | 15 | US-10-087-684-8     | Sequence 8, Appli  |
| 14 | 4488   | 37.4 | 855  | 15 | US-10-218-779-8     | Sequence 8, Appli  |
| 15 | 1902.5 | 15.9 | 401  | 14 | US-10-314-232-15    | Sequence 15, Appl  |
| 16 | 1713.5 | 14.3 | 322  | 14 | US-10-314-232-11    | Sequence 11, Appl  |
| 17 | 1538.5 | 12.8 | 289  | 14 | US-10-314-232-47    | Sequence 47, Appl  |
| 18 | 1327   | 11.1 | 2037 | 15 | US-10-087-684-39    | Sequence 39, Appl  |
| 19 | 1327   | 11.1 | 2037 | 15 | US-10-218-779-39    | Sequence 39, Appl  |
| 20 | 1319   | 11.0 | 2029 | 15 | US-10-087-684-38    | Sequence 38, Appl  |
| 21 | 1319   | 11.0 | 2029 | 15 | US-10-218-779-38    | Sequence 38, Appl  |
| 22 | 1194.5 | 10.0 | 334  | 14 | US-10-314-232-7     | Sequence 7, Appli  |
| 23 | 1188   | 9.9  | 1912 | 17 | US-10-772-636-64    | Sequence 64, Appl  |
| 24 | 1160   | 9.7  | 1767 | 15 | US-10-087-684-40    | Sequence 40, Appl  |
| 25 | 1160   | 9.7  | 1767 | 15 | US-10-218-779-40    | Sequence 40, Appl  |
| 26 | 1158.5 | 9.7  | 1907 | 15 | US-10-291-265-250   | Sequence 250, App  |
| 27 | 1155.5 | 9.6  | 1897 | 17 | US-10-482-029-52    | Sequence 52, Appl  |
| 28 | 1155.5 | 9.6  | 1907 | 18 | US-10-712-892A-34   | Sequence 34, Appl  |
| 29 | 1150   | 9.6  | 1767 | 15 | US-10-087-684-41    | Sequence 41, Appl  |
| 30 | 1150   | 9.6  | 1767 | 15 | US-10-218-779-41    | Sequence 41, Appl  |
| 31 | 1141.5 | 9.5  | 1948 | 9  | US-09-808-602-55    | Sequence 55, Appl  |
| 32 | 1141.5 | 9.5  | 1948 | 10 | US-09-809-198-45    | Sequence 45, Appl  |
| 33 | 1109   | 9.3  | 1997 | 10 | US-09-905-567B-54   | Sequence 54, Appl  |
| 34 | 1108   | 9.2  | 1997 | 16 | US-10-408-765A-2135 | Sequence 2135, Ap  |
| 35 | 1108   | 9.2  | 1997 | 17 | US-10-497-692-4     | Sequence 4, Appli  |
| 36 | 1107   | 9.2  | 1997 | 15 | US-10-634-027-2     | Sequence 2, Appli  |
| 37 | 1026   | 8.6  | 1447 | 17 | US-10-497-692-13    | Sequence 13, Appl  |
| 38 | 966.5  | 8.1  | 1450 | 17 | US-10-497-692-14    | Sequence 14, Appl  |
| 39 | 942    | 7.9  | 176  | 14 | US-10-314-232-9     | Sequence 9, Appli  |
| 40 | 863    | 7.2  | 1238 | 15 | US-10-366-547-47    | Sequence 47, Appl  |
| 41 | 855    | 7.1  | 1367 | 15 | US-10-369-493-5508  | Sequence 5508, Ap  |
| 42 | 855    | 7.1  | 1367 | 15 | US-10-369-493-5509  | Sequence 5509, Ap  |
| 43 | 851.5  | 7.1  | 2041 | 13 | US-10-369-493-5509  | Sequence 5509, Ap  |
| 44 | 839.5  | 7.0  | 1502 | 9  | US-09-808-602-54    | Sequence 54, Appl  |
| 45 | 839.5  | 7.0  | 1502 | 10 | US-09-808-602-54    | Sequence 44, Appl  |

ALIGNMENTS

RESULT 1  
US-09-822-871-2  
; Sequence 2, Application US/09822871  
; Publication No. US20030099942A1  
; GENERAL INFORMATION:  
; APPLICANT: WEBSTER, Marion et al  
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CLO01219  
; CURRENT APPLICATION NUMBER: US/09/822,871  
; CURRENT FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 2291  
; TYPE: PRT  
; ORGANISM: Human  
US-09-822-871-2

|                       |                 |   |           |              |
|-----------------------|-----------------|---|-----------|--------------|
| Query Match           | 100.0%;         | Score 11985;  | DB 10;    | Length 2291; |
| Best Local Similarity | 100.0%;         | Pred. No. 0;  |           |              |
| Matches 2291;         | Conservative 0; | Mismatches 0;   | Indels 0; | Gaps 0;      |
| QY                    | 1               | MDPLIFLLFTGTSETQVDVSNVVGTRYDITISSITTYTSPVTRIVTPNVTKGPPV       | 60        |              |
| Db                    | 1               | MDPLIFLLFTGTSETQVDVSNVVGTRYDITISSITTYTSPVTRIVTPNVTKGPPV       | 60        |              |
| QY                    | 61              | FLAGERVGSAGILLSWNTPPNPENGRIISIVIKYKEVCPMMQTVYTVQVRSPKPSLEVLTN | 120       |              |
| Db                    | 61              | FLAGERVGSAGILLSWNTPPNPENGRIISIVIKYKEVCPMMQTVYTVQVRSPKPSLEVLTN | 120       |              |
| QY                    | 121             | LNPGTTYEIKVAENSAGIGVSDPLPQTASAPGKVNLTVEYNASAVKLIWYLPQ         | 180       |              |

|    |      |     |    |         |         |       |        |       |       |     |      |       |    |    |    |    |     |    |    |    |    |    |    |    |   |     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |   |   |   |   |   |   |     |     |      |      |     |      |     |
|----|------|-----|----|---------|---------|-------|--------|-------|-------|-----|------|-------|----|----|----|----|-----|----|----|----|----|----|----|----|---|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|---|---|---|---|---|---|-----|-----|------|------|-----|------|-----|
| Db | 121  | LNP | GT | TYEIKVA | ENSAGIV | FSDDP | FLFQTA | BSAPG | KVVNL | TV  | EATN | ASAVK | L  | IW | LP | Q  | 180 |    |    |    |    |    |    |    |   |     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |   |   |   |   |   |   |     |     |      |      |     |      |     |
| Qy | 181  | PNG | K  | TSFKIS  | VKH     | ASG   | IV     | KDVS  | IR    | VED | I    | L     | TG | KL | PE | CN | SE  | SE | SL | WT | AS | PS | PL | GR | V | 240 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |   |   |   |   |   |   |     |     |      |      |     |      |     |
| Db | 181  | PNG | K  | TSFKIS  | VKH     | ASG   | IV     | KDVS  | IR    | VED | I    | L     | TG | KL | PE | CN | SE  | SE | SL | WT | AS | PS | PL | GR | V | 240 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |   |   |   |   |   |   |     |     |      |      |     |      |     |
| Qy | 241  | TPS | R  | TH      | SS      | SS    | TL     | QNE   | I     | SS  | V    | KE    | P  | IS | F  | V  | V   | TH | L  | R  | P  | Y  | T  | T  | L | F   | E | V | S | A | A | T | T | E | A | G | I | D | S | T | I | V | R | T | 300 |   |   |   |   |   |   |     |     |      |      |     |      |     |
| Db | 241  | TPS | R  | TH      | SS      | SS    | TL     | QNE   | I     | SS  | V    | KE    | P  | IS | F  | V  | V   | TH | L  | R  | P  | Y  | T  | T  | L | F   | E | V | S | A | A | T | T | E | A | G | I | D | S | T | I | V | R | T | 300 |   |   |   |   |   |   |     |     |      |      |     |      |     |
| Qy | 301  | P   | S  | V       | P       | E     | G      | P     | O     | N   | C    | V     | T  | G  | N  | I  | T   | G  | K  | S  | I  | L  | M  | D  | P | P   | T | I | V | T | G | K | S | I | R | V | E | L | Y | G | P | S | G | R | I   | D | N | S | T | K | D | L   | K   | F    | A    | 360 |      |     |
| Db | 301  | P   | S  | V       | P       | E     | G      | P     | O     | N   | C    | V     | T  | G  | N  | I  | T   | G  | K  | S  | I  | L  | M  | D  | P | P   | T | I | V | T | G | K | S | I | R | V | E | L | Y | G | P | S | G | R | I   | D | N | S | T | K | D | L   | K   | F    | A    | 360 |      |     |
| Qy | 361  | F   | T  | N       | L       | T     | P      | T     | F     | T   | Y    | D     | V  | I  | A  | E  | T   | S  | A  | G  | T  | G  | P  | K  | S | N   | I | S | V | F | T | P | D | V | P | G | A | V | F | O | L | O | L | A | E   | V | E | S | T | O | V | R   | I   | T    | W    | K   | P    | 420 |
| Db | 361  | F   | T  | N       | L       | T     | P      | T     | F     | T   | Y    | D     | V  | I  | A  | E  | T   | S  | A  | G  | T  | G  | P  | K  | S | N   | I | S | V | F | T | P | D | V | P | G | A | V | F | O | L | O | L | A | E   | V | E | S | T | O | V | R   | I   | T    | W    | K   | P    | 420 |
| Qy | 421  | R   | O  | P       | N       | G     | I      | I     | N     | O   | V    | R     | V  | K  | V  | L  | P   | E  | T  | G  | I  | L  | E  | N  | T | L   | L | T | G | N | E | I | N | D | P | M | A | P | E | I | N | I | V | E | P   | M | V | G | L | Y | E | G   | S   | A    | E    | M   | 480  |     |
| Db | 421  | R   | O  | P       | N       | G     | I      | I     | N     | O   | V    | R     | V  | K  | V  | L  | P   | E  | T  | G  | I  | L  | E  | N  | T | L   | L | T | G | N | E | I | N | D | P | M | A | P | E | I | N | I | V | E | P   | M | V | G | L | Y | E | G   | S   | A    | E    | M   | 480  |     |
| Qy | 481  | S   | D  | L       | H       | S     | L      | A     | T     | I   | F    | I     | N  | S  | H  | P  | D   | K  | N  | F  | P  | A  | R  | N  | A | B   | D | O | T | S | P | V | T | T | R | N | O | Y | I | T | D | I | A | E | Q   | L | S | Y | I | R | L | V   | P   | T    | 540  |     |      |     |
| Db | 481  | S   | D  | L       | H       | S     | L      | A     | T     | I   | F    | I     | N  | S  | H  | P  | D   | K  | N  | F  | P  | A  | R  | N  | A | B   | D | O | T | S | P | V | T | T | R | N | O | Y | I | T | D | I | A | E | Q   | L | S | Y | I | R | L | V   | P   | T    | 540  |     |      |     |
| Qy | 541  | E   | H  | M       | I       | S     | V      | A     | T     | I   | M    | G     | E  | G  | P  | T  | V   | L  | S  | V  | R  | T  | R  | O  | O | V   | P | S | S | I | K | I | N | Y | K | N | I | S | S | I | L | L | Y | M | D   | P | P | E | Y | P | N | G   | K   | I    | 600  |     |      |     |
| Db | 541  | E   | H  | M       | I       | S     | V      | A     | T     | I   | M    | G     | E  | G  | P  | T  | V   | L  | S  | V  | R  | T  | R  | O  | O | V   | P | S | S | I | K | I | N | Y | K | N | I | S | S | I | L | L | Y | M | D   | P | P | E | Y | P | N | G   | K   | I    | 600  |     |      |     |
| Qy | 601  | H   | T  | I       | A       | M     | E      | L     | D     | T   | N    | R     | A  | F  | O  | I  | T   | I  | D  | N  | S  | F  | L  | T  | G | L   | K | K | T | Y | K | M | R | V | A | A | S | T | H | D | G | E | S | S | E   | N | D | I | F | V | R | T   | 660 |      |      |     |      |     |
| Db | 601  | H   | T  | I       | A       | M     | E      | L     | D     | T   | N    | R     | A  | F  | O  | I  | T   | I  | D  | N  | S  | F  | L  | T  | G | L   | K | K | T | Y | K | M | R | V | A | A | S | T | H | D | G | E | S | S | E   | N | D | I | F | V | R | T   | 660 |      |      |     |      |     |
| Qy | 661  | S   | E  | D       | E       | P     | S      | S     | Q     | D   | V    | E     | I  | D  | T  | A  | E   | I  | R  | L  | K  | W  | S  | P  | P | E   | K | N | G | I | I | A | E | V | L | Y | K | N | I | D | I | L | Y | M | K   | N | T | S | T | D | I | 720 |     |      |      |     |      |     |
| Db | 661  | S   | E  | D       | E       | P     | S      | S     | Q     | D   | V    | E     | I  | D  | T  | A  | E   | I  | R  | L  | K  | W  | S  | P  | P | E   | K | N | G | I | I | A | E | V | L | Y | K | N | I | D | I | L | Y | M | K   | N | T | S | T | D | I | 720 |     |      |      |     |      |     |
| Qy | 721  | L   | N  | L       | R       | P     | H      | T     | L     | N   | I    | S     | V  | R  | S  | T  | R   | F  | G  | H  | G  | N  | O  | V  | S | L   | L | S | V | R | T | S | E | T | P | D | S | A | P | E | N | I | T | Y | K   | N | I | S | S | G | R | E   | I   | L    | S    | F   | 780  |     |
| Db | 721  | L   | N  | L       | R       | P     | H      | T     | L     | N   | I    | S     | V  | R  | S  | T  | R   | F  | G  | H  | G  | N  | O  | V  | S | L   | L | S | V | R | T | S | E | T | P | D | S | A | P | E | N | I | T | Y | K   | N | I | S | S | G | R | E   | I   | L    | S    | F   | 780  |     |
| Qy | 781  | L   | P  | P       | S       | S     | N      | G     | I     | I   | K    | K     | T  | I  | I  | L  | K   | S  | N  | G  | N  | E  | E  | R  | T | I   | N | T | S | L | T | O | N | I | K | V | L | K | K | T | O | Y | I | I | E   | V | S | A | S | T | L | K   | G   | E    | V    | 840 |      |     |
| Db | 781  | L   | P  | P       | S       | S     | N      | G     | I     | I   | K    | K     | T  | I  | I  | L  | K   | S  | N  | G  | N  | E  | E  | R  | T | I   | N | T | S | L | T | O | N | I | K | V | L | K | K | T | O | Y | I | I | E   | V | S | A | S | T | L | K   | G   | E    | V    | 840 |      |     |
| Qy | 841  | R   | S  | A       | P       | I     | S      | I     | L     | T   | E    | A     | D  | S  | P  | P  | O   | D  | S  | V  | K  | L  | S  | G  | V | T   | V | L | S | W | O | P | P | L | E | P | N | G | I | I | L | Y | T | T | Y   | V | V | N | R | S | S | L   | K   | T    | 900  |     |      |     |
| Db | 841  | R   | S  | A       | P       | I     | S      | I     | L     | T   | E    | A     | D  | S  | P  | P  | O   | D  | S  | V  | K  | L  | S  | G  | V | T   | V | L | S | W | O | P | P | L | E | P | N | G | I | I | L | Y | T | T | Y   | V | V | N | R | S | S | L   | K   | T    | 900  |     |      |     |
| Qy | 901  | I   | N  | V       | T       | E     | S      | L     | E     | S   | L    | D     | L  | D  | Y  | N  | V   | E  | S  | A  | Y  | T  | A  | S  | T | R   | F | G | D | G | T | G | S | N | I | I | S | F | O | T | P | E | G | A | P   | S | D | P | P | K | O | V   | Y   | A    | N    | L   | 960  |     |
| Db | 901  | I   | N  | V       | T       | E     | S      | L     | E     | S   | L    | D     | L  | D  | Y  | N  | V   | E  | S  | A  | Y  | T  | A  | S  | T | R   | F | G | D | G | T | G | S | N | I | I | S | F | O | T | P | E | G | A | P   | S | D | P | P | K | O | V   | Y   | A    | N    | L   | 960  |     |
| Qy | 961  | S   | S  | S       | I       | L     | F      | T     | P     | P   | S    | K     | P  | N  | G  | I  | I   | Q  | Y  | S  | V  | Y  | R  | N  | T | S   | G | T | P | M | O | N | F | T | L | H | E | L | N | D | N | D | M | T | V   | S | T | I | D | I | K | L   | T   | 1020 |      |     |      |     |
| Db | 961  | S   | S  | S       | I       | L     | F      | T     | P     | P   | S    | K     | P  | N  | G  | I  | I   | Q  | Y  | S  | V  | Y  | R  | N  | T | S   | G | T | P | M | O | N | F | T | L | H | E | L | N | D | N | D | M | T | V   | S | T | I | D | I | K | L   | T   | 1020 |      |     |      |     |
| Qy | 1021 | F   | S  | Y       | T       | T     | F      | L     | T     | A   | S    | T     | S  | V  | G  | N  | K   | S  | S  | D  | I  | I  | E  | V  | Y | T   | D | O | I | P | E | G | F | V | G | N | L | T | Y | E | S | I | S | T | A   | I | N | S | W | P | P | A   | Q   | P    | 1080 |     |      |     |
| Db | 1021 | F   | S  | Y       | T       | T     | F      | L     | T     | A   | S    | T     | S  | V  | G  | N  | K   | S  | S  | D  | I  | I  | E  | V  | Y | T   | D | O | I | P | E | G | F | V | G | N | L | T | Y | E | S | I | S | T | A   | I | N | S | W | P | P | A   | Q   | P    | 1080 |     |      |     |
| Qy | 1081 | N   | G  | L       | V       | F     | Y      | V     | S     | L   | I    | L     | O  | O  | T  | P  | R   | H  | R  | P  | L  | V  | T  | T  | E | R   | S | I | Y | F | D | N | L | E | K | Y | T | D | I | L | K | I | T | P | S   | T | E | K | G | S | D | P   | T   | Y    | A    | Q   | 1140 |     |
| Db | 1081 | N   | G  | L       | V       | F     | Y      | V     | S     | L   | I    | L     | O  | O  | T  | P  | R   | H  | R  | P  | L  | V  | T  | T  | E | R   | S | I | Y | F | D | N | L | E | K | Y | T | D | I | L | K | I | T | P | S   | T | E | K | G | S | D | P   | T   | Y    | A    | Q   | 1140 |     |
| Qy | 1141 | L   | Y  | I       | K       | T     | E      | D     | P     | E   | T    | S     | P  | I  | N  | T  | F   | K  | N  | L  | S  | T  | S  | V  | L | S   | W | D | P | V | K | P | N | G | A | I | S | Y | D | L | T | L | O | G | P   | N | E | N | S | F | I | T   | S   | 1200 |      |     |      |     |
| Db | 1141 | L   | Y  | I       | K       | T     | E      | D     | P     | E   | T    | S     | P  | I  | N  | T  | F   | K  | N  | L  | S  | T  | S  | V  | L | S   | W | D | P | V | K | P | N | G | A | I | S | Y | D | L | T | L | O | G | P   | N | E | N | S | F | I | T   | S   | 1200 |      |     |      |     |
| Qy | 1201 | D   | N  | Y       | I       | L     | E      | E     | L     | S   | P    | F     | T  | L  | Y  | S  | F   | F  | A  | A  | R  | T  | K  | G  | L | P   | S | S | I | L | F | F | Y | T | D | S | V | P | L | A | P | P | O | N | L   | T | L | I | N | C | T | S   | D   | F    | V    | M   | 1260 |     |
| Db | 1201 | D   | N  | Y       | I       | L     | E      | E     | L     | S   | P    | F     | T  | L  | Y  | S  | F   | F  | A  | A  | R  | T  | K  | G  | L | P   | S | S | I | L | F | F | Y | T | D | S | V | P | L | A | P | P | O | N | L   | T | L | I | N | C | T | S   | D   | F    | V    | M   | 1260 |     |

RESULT 2  
US-10-673-885-2  
; Sequence 2, Application US/10673885  
; Publication No. US20040081644A1  
; GENERAL INFORMATION:  
; APPLICANT: BEASLEY, Ellen et al  
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CL001219D1V  
; CURRENT APPLICATION NUMBER: US/10/673,885  
; CURRENT FILING DATE: 2003-09-30  
; PRIOR APPLICATION NUMBER: 09/822,871  
; PRIOR FILING DATE: 04-02-2001  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 2291  
; TYPE: PRP  
; ORGANISM: Human  
US-10-673-885-2

Query Match 100.0%; Score 11985; DB 15; Length 2291;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MDPLIIFLLFIQTSQDVSNVPCRYDITISISITSTYTSPTVTRIVTPNVTKGPPV 60  
DB 1 MDPLIIFLLFIQTSQDVSNVPCRYDITISISITSTYTSPTVTRIVTPNVTKGPPV 60  
  
QY 61 FLA GERVSGAGILLSNWTPNPNNGRIISYIVKVEKCPNMQTYTVQSRKPSLEVLTN 120  
DB 61 FLA GERVSGAGILLSNWTPNPNNGRIISYIVKVEKCPNMQTYTVQSRKPSLEVLTN 120  
  
QY 121 LNPCTTYEIKVAENSGAGVFDPLQFAESAPGVNLTVEAYNASAVKLIWYLPQ 180  
DB 121 LNPCTTYEIKVAENSGAGVFDPLQFAESAPGVNLTVEAYNASAVKLIWYLPQ 180  
  
QY 181 PCKITSFKISVKAHSGIVKDVSRVEDILGKLPCECNENSEFLWSTASPSPTLGRV 240  
DB 181 PCKITSFKISVKAHSGIVKDVSRVEDILGKLPCECNENSEFLWSTASPSPTLGRV 240  
  
QY 241 TPSSRTTHSSSTLTQNEISSVWKEPISFVVTTLRPVTTTLFEVSAATTEAGYIDSTIVRT 300  
DB 241 TPSSRTTHSSSTLTQNEISSVWKEPISFVVTTLRPVTTTLFEVSAATTEAGYIDSTIVRT 300  
  
QY 301 PESVPGPPQNCVTGNITGKSFILWDPPTIVTGKFSYRVELYGSGRILDNSTKDKFA 360  
DB 301 PESVPGPPQNCVTGNITGKSFILWDPPTIVTGKFSYRVELYGSGRILDNSTKDKFA 360  
  
QY 361 FTNLTPTTYDVYIAAETSAGTCKSNISVFTPPDPVGA VFDLQLAEVSTOVRITWKPK 420  
DB 361 FTNLTPTTYDVYIAAETSAGTCKSNISVFTPPDPVGA VFDLQLAEVSTOVRITWKPK 420  
  
QY 421 RQPNGIINQYRVKVLVPETGIIILENTLLTGNNYEIINDPMAPEIVNIVEPMVGLYEGSAEM 480  
DB 421 RQPNGIINQYRVKVLVPETGIIILENTLLTGNNYEIINDPMAPEIVNIVEPMVGLYEGSAEM 480  
  
QY 481 SSDLSLATPIYNHSDPKPNPARRAEDQTSVVTTRNQYITDIAEQLSYVIRRLVPPT 540  
DB 481 SSDLSLATPIYNHSDPKPNPARRAEDQTSVVTTRNQYITDIAEQLSYVIRRLVPPT 540  
  
QY 541 EHMISVSAPFTIMEGEGPTVLVSRTQOVPSIKIINVKNISSSSILLYWDPPPEYNGKIT 600  
DB 541 EHMISVSAPFTIMEGEGPTVLVSRTQOVPSIKIINVKNISSSSILLYWDPPPEYNGKIT 600  
  
QY 601 HYTIYAMELDTRAFQITIDNSFLITGLKKYTKYKRWVAASTHDGESSISENDIFVRT 660  
DB 601 HYTIYAMELDTRAFQITIDNSFLITGLKKYTKYKRWVAASTHDGESSISENDIFVRT 660  
  
QY 661 SEDEPSSQDVVEIDVTADIEIRLKWSPPEKNGIIIAVEVLKYNIDTLYMKNTSTTDII 720  
DB 661 SEDEPSSQDVVEIDVTADIEIRLKWSPPEKNGIIIAVEVLKYNIDTLYMKNTSTTDII 720

661 SEDEPSSQDVVEIDVTADIEIRLKWSPPEKNGIIIAVEVLKYNIDTLYMKNTSTTDII 720  
721 LRNLRPHTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSDAPENITYKNISSEIELSP 780  
721 LRNLRPHTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSDAPENITYKNISSEIELSP 780  
781 LPPSSPNGIIKKYTVILKESNGNEERTINTTSITONIKVLKKYTOVILIEVSASTLKGEV 840  
781 LPPSSPNGIIKKYTVILKESNGNEERTINTTSITONIKVLKKYTOVILIEVSASTLKGEV 840  
841 RSAPISILTEEDAPSPQDQFSVKQLSGVTVKLSWQPLEPNGIILYYTVVWNRSLKT 900  
841 RSAPISILTEEDAPSPQDQFSVKQLSGVTVKLSWQPLEPNGIILYYTVVWNRSLKT 900  
901 INVTETSLSLDLNVDNYSAYVTASTRFGDGKGTGSIISFQPEGAPSDPPKDVVYANL 960  
901 INVTETSLSLDLNVDNYSAYVTASTRFGDGKGTGSIISFQPEGAPSDPPKDVVYANL 960  
961 SSSSIILFWTPPSKNGIIQYYSVYRNTSGTFPMQNTLHFLTNDNDNMTVSTIIDKLTII 1020  
961 SSSSIILFWTPPSKNGIIQYYSVYRNTSGTFPMQNTLHFLTNDNDNMTVSTIIDKLTII 1020  
1021 FSYTFLWTASTSVGNGKSSDIIIEVYTDQDIPGFGVGNLTYESISSTAINVSWVPAQP 1080  
1021 FSYTFLWTASTSVGNGKSSDIIIEVYTDQDIPGFGVGNLTYESISSTAINVSWVPAQP 1080  
1081 NGLVFFYVSLIILOQTPRHVRPPLVTVYERSIYFDNLEKYTDYILKITPSTEGFSDVYTAQ 1140  
1081 NGLVFFYVSLIILOQTPRHVRPPLVTVYERSIYFDNLEKYTDYILKITPSTEGFSDVYTAQ 1140  
1141 LYKTEEDVPETSPIINTFNLSSTSVLLSWDPVPKPGALISYDYLTLQGNENYSPITS 1200  
1141 LYKTEEDVPETSPIINTFNLSSTSVLLSWDPVPKPGALISYDYLTLQGNENYSPITS 1200  
1201 DNVIIILEUSPFTLSFFAAARTRKGLGPSSILFFYTDSEVPLAPQNLTLINCTSDFW 1260  
1201 DNVIIILEUSPFTLSFFAAARTRKGLGPSSILFFYTDSEVPLAPQNLTLINCTSDFW 1260  
1261 LKMSPLPGGIVKVSFKIHEHETDIYKXISGFKTEAKLVGLPEVSTYSIRVSAFTK 1320  
1261 LKMSPLPGGIVKVSFKIHEHETDIYKXISGFKTEAKLVGLPEVSTYSIRVSAFTK 1320  
1321 VGNQFQSNVVKFTQESVDPVQNMCMATSNQSVLVKWDPPKANGIIITQMTVVERN 1380  
1321 VGNQFQSNVVKFTQESVDPVQNMCMATSNQSVLVKWDPPKANGIIITQMTVVERN 1380  
1381 STKVSQDQHWYTFIKLLANTSVYKVRASTSAGEGDESTCHVSTLTPETVPSVPTNIAFSD 1440  
1381 STKVSQDQHWYTFIKLLANTSVYKVRASTSAGEGDESTCHVSTLTPETVPSVPTNIAFSD 1440  
1441 VQSTSATLTWIRPDITLGYFQNYKITTLQRAQCKEWESEECVEYQKIQVLYEAHLTEET 1500  
1441 VQSTSATLTWIRPDITLGYFQNYKITTLQRAQCKEWESEECVEYQKIQVLYEAHLTEET 1500  
1501 VYGLKKFRWYRFQVAASTNAGYGNASNWTSTKTLPGPPDGPENHVHVAATSPSISWS 1560  
1501 VYGLKKFRWYRFQVAASTNAGYGNASNWTSTKTLPGPPDGPENHVHVAATSPSISWS 1560  
1561 EBAVITGPTCYLIDVKSVDNDEFNISFIKSNENKTIETIKLEIFRYSVWITAFNGNIS 1620  
1561 EBAVITGPTCYLIDVKSVDNDEFNISFIKSNENKTIETIKLEIFRYSVWITAFNGNIS 1620  
1621 AAYVEGKSAEMIVTTLSAPKDPNNMTFKIPDBVTQFQLTFLPSPNGNIQVYQAL 1680  
1621 AAYVEGKSAEMIVTTLSAPKDPNNMTFKIPDBVTQFQLTFLPSPNGNIQVYQAL 1680  
1681 VYREDDPTAVQIHNLSIIQKTNFTVIAMLEGLKGGHTYNI SVYVANSAGAGPKVPMRITM 1740  
1681 VYREDDPTAVQIHNLSIIQKTNFTVIAMLEGLKGGHTYNI SVYVANSAGAGPKVPMRITM 1740  
1741 DIKAPARPKTKPTPIYDATGKLLVSTTTITRMPICYSDDHGPIKNVQVLAETGCAQHD 1800  
1741 DIKAPARPKTKPTPIYDATGKLLVSTTTITRMPICYSDDHGPIKNVQVLAETGCAQHD 1800

```
QY 1801 GNVTKWYDAYFNKARPYFTNEGPNPCTEGKTKFSGNEEIIYIIGADNACMIPEGNEDKIC 1860
Db 1801 GNVTKWYDAYFNKARPYFTNEGPNPCTEGKTKFSGNEEIIYIIGADNACMIPEGNEDKIC 1860
QY 1861 NGPLKPKQYLFKFRATNIMGQFTDSYSDPVKTLGEGLSERTVEIILSVTLCILSIILL 1920
Db 1861 NGPLKPKQYLFKFRATNIMGQFTDSYSDPVKTLGEGLSERTVEIILSVTLCILSIILL 1920
QY 1921 GTAIFAFARIRKQKEGGTYSQDAEIIIDTKLKLQDLITVADLEKDERLTRPIKSKSFL 1980
Db 1921 GTAIFAFARIRKQKEGGTYSQDAEIIIDTKLKLQDLITVADLEKDERLTRPIKSKSFL 1980
QY 1981 QHVEELCTNNLKFQEFSELKFLQDLISDADLPWNAKRFNPKYNNNNRKLIIA 2040
Db 1981 QHVEELCTNNLKFQEFSELKFLQDLISDADLPWNAKRFNPKYNNNNRKLIIA 2040
QY 2041 DASVPGSDVINASYISGYLCPNEFIATQGPLGTVGDFWRMVWETRAKTLVMLTQCFEKG 2100
Db 2041 DASVPGSDVINASYISGYLCPNEFIATQGPLGTVGDFWRMVWETRAKTLVMLTQCFEKG 2100
QY 2101 RIRCHQWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVRCNFTAWPE 2160
Db 2101 RIRCHQWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVRCNFTAWPE 2160
QY 2161 HGVPEASAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGVFTALDHLTOHINDHDFVDI 2220
Db 2161 HGVPEASAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGVFTALDHLTOHINDHDFVDI 2220
QY 2221 YGLVABLRSERMCWONLAQYIFLHOCILDLNKGNSNOPICFVNYSAIQKQMSLDAMEG 2280
Db 2221 YGLVABLRSERMCWONLAQYIFLHOCILDLNKGNSNOPICFVNYSAIQKQMSLDAMEG 2280
QY 2281 DVELEWEETM 2291
Db 2281 DVELEWEETM 2291
```

```
RESULT 3
US-10-087-684-10
; Sequence 10, Application US/10087684
; Publication No. US20040029116A1
; GENERAL INFORMATION:
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: MacDougall, John R.
; APPLICANT: Millet, Isabelle
; APPLICANT: Ellerman, Karen
; APPLICANT: Stone, David J.
; APPLICANT: Grose, William M.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie, J.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Mishra, Vishnu
; APPLICANT: Sheno, Suresh G.
; APPLICANT: Rastelli, Luca
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Miller, Charles E.
; APPLICANT: Gangolli, Esha A.
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-214 CIP
; CURRENT APPLICATION NUMBER: US/10/087,684
; CURRENT FILING DATE: 2003-03-10
; PRIOR APPLICATION NUMBER: 60/253,834
; PRIOR FILING DATE: 2000-11-29
```

```
; PRIOR APPLICATION NUMBER: 60/250,926
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/264,180
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,656
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/327,456
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 10
; LENGTH: 2300
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-087-684-10
```

```
Query Match 98.5%; Score 11807.5; DB 15; Length 2300;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 2276; Conservative 5; Mismatches 5; Indels 19; Gaps 9;

QY 1 MDLFIILLFLLFISGTSETQVDVSNVVPGRYDITISSITSTYSPVTRIVTPNVTGPPVP 60
Db 1 MDLFIILLFLLFISGTSETQVDVSNVVPGRYDITISSITSTYSPVTRIVTPNVTGPPVP 60
QY 61 FLAGERVGSAGILLSWNTPPNPNGRILSIYIVKYKEVCPMMQVTVTVQVRSKPSLEVLITN 120
Db 61 FLAGERVGSAGILLSWNTPPNPNGRILSIYIVKYKEVCPMMQVTVTVQVRSKPSLEVLITN 120
QY 121 LNPQTTYEIKVAENSAGIGVSPDPLFQTAES-APGKVNLTVZAYNASAVKLIWYLP 179
Db 121 LNPQTTYEIKVAENSAGIGVSPDPLFQTAESAPGKVNLTVZAYNASAVKLIWYLP 180
QY 180 QPNGKITSFKISVKHARSIGVVKDYSIRVEDILTGLPECN-ENSEFSLWSTASPSPTLG 238
Db 181 QPNGKITSFKISVKHARSIGVVKDYSIRVEDILTGLPECNVENSEFSLWSTASPSPTLG 240
QY 239 RVTPPSRTHSSSTLTQNEISSWKEPISFVVTHLRPVTYTLFEVSAATTEAGYIDSTIV 298
Db 241 RVTPPSRTHSSSTLTQNEISSWKEPISFVVTHLRPVTYTLFEVSAATTEAGYIDSTIV 300
QY 299 RTPESVPEGPONCVTGNITGKSFILWDPPTIVTGKSYRVELYGPSGRILDNSTKOLK 358
Db 301 RTPESVPEGPONCVTGNITGKSFILWDPPTIVTGKSYRVELYGPSGRILDNSTKOLK 360
QY 359 FAPTNLTPPTMYDVYIAAETSAGTGPKSNISVFTPPDVPDGLQLAESTQVRITWK 418
Db 361 FAPTNLTPPTMYDVYIAAETSAGTGPKSNISVFTPPDVPDGLQLAESTQVRITWK 420
QY 419 KRPQNGIINQVRKVLVPETGIIILENTLLTGNNEYINDPMAPEIINIVEPMVGLYEGSA 478
Db 421 KRPQNGIINQVRKVLVPETGIIILENTLLTGNNE-INDPMAPEIINIVEPMVGLYEGSA 479
QY 479 EMSSDLHSLATFIYNSHDPKPNARNABDQTSPPVVTTRNOYITDIAAQLSVYRRLVP 538
Db 480 EMSSDLHSLATFIYNSHDPKPNARNABDQTSPPVVTTRNOYITDIAAQLSVYRRLVP 539
QY 539 FTEHMISVSAFTIMGEGPPTVLVSRTRQOVPSIKIINYKNISSSILLYWDPPEYPNK 598
Db 540 FTEHMISVSAFTIMGEGPPTVLVSRTRQOVPSIKIINYKNISSSILLYWDPPEYPNK 599
QY 599 ITHYTIYAMELDTNRAFOITTTIDNSFLIT--GLKYYTKYKRVAASTHDSSESLSENDI 656
Db 600 ITHYTIYAMELDTNRAFOITTTIDNSFLITIGLKKYTKYKRVAASTHDSSESLSENDI 659
QY 657 FVRTSEDBESSPOQVEVIDVTADIRLKWSPPEKPNGIIIAYEVLVYKNIDTLYMKNVST 716
Db 660 FVRTSEDBESSPOQVEVIDVTADIRLKWSPPEKPNGIIIAYEVLVYKNIDTLYMKNVST 719
QY 717 TDIIILNLRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSGEI 776
Db 720 TDIIILNLRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSGEI 779
```

QY 777 ELSFLPPSPNGIIKKYIYVLRKSGNGBERTINTSLTONIKVLKKYTOVIIIVSASTLK 836  
DB 780 ELSFLPPSPNGIIKKYIYVLRKSGNGBERTINTSLTONI--LKKYTOVIIIVSASTLK 837  
QY 837 GEGVRSAPISILTEEDAPSPPODFSVKQVSGVTVKLSWQPPLEPNGLIILYTVVYVW-NR 895  
DB 838 GEGVRSAPISILTEEDAPSPPODFSVKQVSGVTVKLSWQPPLEPNGLIILYTVVYVWR 897  
QY 896 SSLKTIINVTETSLSDLDYVVEYSAYVASTASTFPGDKTGSNIISFQTEGAPSDPPKDV 955  
DB 898 SSLKTIINVTETSLSDLDYVVEYSAYVASTASTFPGDKTGSNIISFQTEG-PSDPPKDV 956  
QY 956 YYANLSSSSIIILFWTPSPKNGIIQYYSVYVYRNTSGTFMONTFLHELTNDNDNMTVSTII 1015  
DB 957 YYANLSSSSIIILFWTPSPKNGIIQYYSVYVYRNTSGTFMONTFLHELTNDNDNMTVSTII 1016  
QY 1016 DKLTIFSYTFLWTASTSVGNKSSDIIEVYTDQDIPGFGVGNLTYESISSTAINVSW 1075  
DB 1017 DKLTIFSYTFLWTASTSVGNKSSDIIEVYTDQDIPGFGVGNLTYESISSTAINVSW 1076  
QY 1076 PPAQPNGLVYVYVSLIILOTPRHVRPPLVYERSIYFDNLEKTYDILKLTPTSTKGFSD 1135  
DB 1077 PPAQPNGLVYVYVSLIILOTPRHVRPPLVYERSIYFDNLEKTYDILKLTPTSTKGFSD 1136  
QY 1136 TYTAQYIKTEEDVPETSPINTPKNLSSTSVLLSWDPPVKPNGAIIISYDLTLQGNENY 1195  
DB 1137 TYTAQYIKTEEDVPETSPINTPKNLSSTSVLLSWDPPVKPNGAIIISYDLTLQGNENY 1196  
QY 1196 SPITSDNYIILBLSPTLYSFAAARTKGLGSPSILPYTDESVPPLAPQNLTLINCT 1255  
DB 1197 SPITSDNYIILBLSPTLYSFAAARTKGLGSPSILPYTDESVPPLAPQNLTLINCT 1256  
QY 1256 SDFVWLKWSPLPGGIVKYVSKIEHEHETDTIYKNIISGFKTEAKLVLEPVSTYSIRV 1315  
DB 1257 SDFVWLKWSPLPGGIVKYVSKIEHEHETDTIYKNIISGFKTEAKLVLEPVSTYSIRV 1316  
QY 1316 SAFTKVGNGQNFVNKFTTQESVPDVQVQNMCMATSWQSVLVKWDPPKKKANGIIITQYVW 1375  
DB 1317 SAFTKVGNGQNFVNKFTTQESVPDVQVQNMCMATSWQSVLVKWDPPKKKANGIIITQYVW 1376  
QY 1376 TVERNSTKVSQDHHMTFFIKLLANTSVYFKVRASTAGEGDESTCHVSTLPETVPSVPTN 1435  
DB 1377 TVERNSTKVSQDHHMTFFIKLLANTSVYFKVRASTAGEGDESTCHVSTLPETVPSVPTN 1436  
QY 1436 IAFSDVQSTSATLTWIRPDITLGFQNYKITTLQRAQCKEWESECEVQKIQYLYEAH 1495  
DB 1437 IAFSDVQSTSATLTWIRPDITLGFQNYKITTLQRAQCKEWESECEVQKIQYLYEAH 1496  
QY 1496 LTBETVYGLKFRWYRFQVAASNAAGVGNASNMISTKTLPGPPDGPENNVHVVATSPFSI 1555  
DB 1497 LTBETVYGLKFRWYRFQVAASNAAGVGNASNMISTKTLPGPPDGPENNVHVVATSPFSI 1556  
QY 1556 SISWSBPAVITGTCYLDIVKSVNDNEFNISPTIKSNEENKTIBIKOLEIFTRYSVVITAF 1615  
DB 1557 SISWSBPAVITGTCYLDIVKSVNDNEFNISPTIKSNEENKTIBIKOLEIFTRYSVVITAF 1616  
QY 1616 TGNISAAVEGKSSAEMIVTLESAPKDPNNMTFQKIDPDEVTKFQTLPPSPQNGNQ 1675  
DB 1617 TGNISAAVEGKSSAEMIVTLESAPKDPNNMTFQKIDPDEVTKFQTLPPSPQNGNQ 1676  
QY 1676 VYQALVYREDDTAVQIHNLSIIQKNTFVIAMLEGLKGHTYNIISVYVNSAGAGPKVP 1735  
DB 1677 VYQALVYREDDTAVQIHNLSIIQKNTFVIAMLEGLKGHTYNIISVYVNSAGAGPKVP 1736  
QY 1736 MRITMDIKAPARKTPTPIYDATGKLLVSTTTITIRMPICYSDHGHPIKNVQVLATET 1795  
DB 1737 MRITMDIKAPARKTPTPIYDATGKLLVSTTTITIRMPICYSDHGHPIKNVQVLATET 1796  
QY 1796 GAQHDGNVTKWYDAYFNKARPYTNEGFPNPPCTEGKTKFSGNEEYIIGADNACMIPGN 1855  
DB 1797 GAQHDGNVTKWYDAYFNKARPYTNEGFPNPPCTEGKTKFSGNEEYIIGADNACMIPGN 1856

QY 1856 EDKICNGPLKPKQYLFKERATNIMQFTDSYSDPVKTLGEGLSERTVEIILSVTLCIL 1915  
DB 1857 EDKICNGPLKPKQYLFKERATNIMQFTDSYSDPVKTLGEGLSERTVEIILSVTLCIL 1916  
QY 1916 SIILIGTALFAPARIRQKQEGTYSPODAEIIDTKLQDQITVADLELKDBRLTR--- 1972  
DB 1917 SIILIGTALFAPARIRQKQEGTYSPODAEIIDTKLQDQITVADLELKDBRLTRLLS 1976  
QY 1973 -----PISKKSFLQHVLEELCTNNNLKQBEFSELPKFQDLSSTADLPWNRKRNFPN 2026  
DB 1977 YRKSIIKPIKSKSFLQHVLEELCTNNNLKQBEFSELPKFQDLSSTADLPWNRKRNFPN 2036  
QY 2027 IKPYNNNRVKLIADASVPGSDIYNASISYGLCPNEFTATQGPLCTGTFDPRWYMETR 2086  
DB 2037 IKPY-NNNRVKLIADASVPGSDIYNASISYGLCPNEFTATQGPLCTGTFDPRWYMETR 2095  
QY 2087 AKTLVMTQCFEGRIRCHQYWPEDNKPTVFGDIVITKLMEDVQIDWTIRDLKIERHGD 2146  
DB 2096 AKTLVMTQCFEGRIRCHQYWPEDNKPTVFGDIVITKLMEDVQIDWTIRDLKIERHGD 2155  
QY 2147 CMTVRQCNFTAPWPEHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVFIALD 2206  
DB 2156 CMTVRQCNFTAPWPEHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVFIALD 2215  
QY 2207 HUTOHINDHDFVDIYGLVAELSRMCMQNLQAIYIFLHCQILDLLSNKSGNQPICFVNY 2266  
DB 2216 HUTOHINDHDFVDIYGLVAELSRMCMQNLQAIYIFLHCQILDLLSNKSGNQPICFVNY 2275  
QY 2267 SALQKMDSLDAMEGDGVELEWEETTM 2291  
DB 2276 SALQKMDSLDAMEGDGVELEWEETTM 2300

## RESULT 4

US-10-218-779-10

; Sequence 10, Application US/10218779

; Publication No. US20040029222A1

; GENERAL INFORMATION:

; APPLICANT: Edinger, Shlomit

; APPLICANT: MacDougall, John

; APPLICANT: Millett, Isabelle

; APPLICANT: Ellerman, Karen

; APPLICANT: Stone, David

; APPLICANT: Gerlach, Valerie

; APPLICANT: Grosse, William

; APPLICANT: Alsobrook II, John

; APPLICANT: Lepley, Denise

; APPLICANT: Rieger, Daniel

; APPLICANT: Burgess, Catherine

; APPLICANT: Casman, Stacie

; APPLICANT: Spytek, Kimberly

; APPLICANT: Boldog, Ferenc

; APPLICANT: Li, Li

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Mishra, Vishnu

; APPLICANT: Patturajan, Meera

; APPLICANT: Shenoy, Suresh

; APPLICANT: Tchernev, Velizar

; APPLICANT: Vernet, Corine

; APPLICANT: Zerhusen, Bryan

; APPLICANT: Malyankar, Uriel

; APPLICANT: Guo, Xiaojia

; APPLICANT: Miller, Charles

; APPLICANT: Gangolli, Esha

; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

; FILE REFERENCE: 21402-214

; CURRENT APPLICATION NUMBER: US/10/218,779

; CURRENT FILING DATE: 2002-08-14

; PRIOR APPLICATION NUMBER: 60/253,834

; PRIOR FILING DATE: 2000-11-29

; PRIOR APPLICATION NUMBER: 60/250,-926

; PRIOR FILING DATE: 2000-11-30

;; PRIOR APPLICATION NUMBER: 60/264,180  
;; PRIOR FILING DATE: 2001-01-25  
;; PRIOR APPLICATION NUMBER: 60/313,656  
;; PRIOR FILING DATE: 2001-08-20  
;; PRIOR APPLICATION NUMBER: 60/327,456  
;; PRIOR FILING DATE: 2001-10-05  
;; NUMBER OF SEQ ID NOS: 216  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 10  
;; LENGTH: 2300  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-218-779-10

Query Match 98.5%; Score 11807.5; DB 15; Length 2300;  
Best Local Similarity 98.7%; Pred. No. 0;  
Matches 2276; Conservative 5; Mismatches 5; Indels 19; Gaps 9;  
QY 1 MDPLIIFLLFLLTGTSETQVDNSVNVPGTRYDITISISITTYTSPVTRIVTPNVTKPGPV 60  
DB 1 MDPLIIFLLFLLTGTSETQVDNSVNVPGTRYDITISISITTYTSPVTRIVTPNVTEGPPV 60  
QY 61 FLAGERVGSAGILLSNWTPPNNGRIISYIVKYKEVCPMMQTVYTVQVRKSPDSLEVLTN 120  
DB 61 FLAGERVGSAGILLSNWTPPNNGRIISYIVKYKEVCPMMQTVYTVQVRKSPDSLEVLTN 120  
QY 121 INPGITYEIKVAENSAGIGVSDPFLPQTAES-APGKVNLTVAYNASAVKLIWYLR 179  
DB 121 INPGITYEIKVAENSAGIGVSDPFLPQTAES-APGKVNLTVAYNASAVKLIWYLR 180  
QY 180 QPQNGKITSPKISVGHARSIGVVDVSVIRVEDILTGKLPKCN-ENSESFLWSTASPSPTLG 238  
DB 181 QPQNGKITSPKISVGHARSIGVVDVSVIRVEDILTGKLPKCNVENSESFLWSTASPSPTLG 240  
QY 239 RVTTPSRTHSSSTLTQNEBISVWKEPISFVTVHLRPYTYTLFEVSAATTEAGYIDSTIV 298  
DB 241 RVTTPSRTHSSSTLTQNEBISVWKEPISFVTVHLRPYTYTLFEVSAATTEAGYIDSTIV 300  
QY 299 RTPESVPEGPONCVTGNITGKSFSLMDPPTIVTGKFSYRVELYGPSRIILDNSTKDLK 358  
DB 301 RTPESVPEGPONCVTGNITGKSFSLMDPPTIVTGKFSYRVELYGPSRIILDNSTKDLK 360  
QY 359 FAFNTLTPPTMDVYIAASTAGTGPKNSISVFTPDVPGAVFDLQLAVESTQVRIWTK 418  
DB 361 FAFNTLTPPTMDVYIAASTAGTGPKNSISVFTPDVPGAVFDLQLAVESTQVRIWTK 420  
QY 419 KPRQNGIINQYRVKLVVPETGIIILENTLLTGNEYINDPMAPEIVNIIVEPMVGLYEGSA 478  
DB 421 KPRQNGIINQYRVKLVVPETGIIILENTLLTGNEB-INDPMAPEIVNIIVEPMVGLYEGSA 479  
QY 479 EMSSDLHSLATFIYNSHPDKNFPARNRABDQTSVVTTRNQVITDIAABQLSYVIRRLVP 538  
DB 480 EMSSDLHSLATFIYNSHPDKNFPARNRABDQTSVVTTRNQVITDIAABQLSYVIRRLVP 539  
QY 539 FTEHMSVSAFTIMGEGPPTVLSVTRQOVPSIKIINYKNISSSILLYMDPPYPNCK 598  
DB 540 FTEHMSVSAFTIMGEGPPTVLSVTRQOVPSIKIINYKNISSSILLYMDPPYPNCK 599  
QY 599 ITHYTIYAMELDTNRAFOITTDNSPLIT--GLKKYTKYKMRVAASTHDSGSSLEENDI 656  
DB 600 ITHYTIYAMELDTNRAFOITTDNSPLITGIGLKKYTKYKMRVAASTHDSGSSLEENDI 659  
QY 657 FVRTSEDESPQDVEVIDVTADEIRLWSPPEKPGNGIIAYEVLVYKNIIDTLYMKNTST 716  
DB 660 FVRTSEDESPQDVEVIDVTADEIRLWSPPEKPGNGIIAYEVLVYKNIIDTLYMKNTST 719  
QY 717 TDIILRLNLPHTLYNLSVSRYTFGHNQVSSILLSVURTSETVPSAPENITVKNISSGRI 776  
DB 720 TDIILRLNLPHTLYNLSVSRYTFGHNQVSSILLSVURTSETVPSAPENITVKNISSGRI 779  
QY 777 ELSFLPPSPNGIISKYTYLKRSGNEERTINTTSLTONIKVLKXYQYIIIEVSASTLK 836  
DB 780 ELSFLPPSPNGIISKYTYLKRSGNEERTINTTSLTONI--LKKYTYIIIEVSASTLK 837

QY 837 GEGVRSAPISILTEEDADPSPQDPFSVKQLSGVTVKLSWQPPLEPNGLIYYTVVW-NR 895  
DB 838 GEGVRSAPISILTEEDADPSPQDPFSVKQLSGVTVKLSWQPPLEPNGLIYYTVVWNR 897  
QY 896 SSLKTIINVTETSLSDLDYVVEYSAYVTASTRFGDGKTSNIIISFOTPEGAPSPKDV 955  
DB 898 SSLKTIINVTETSLSDLDYVVEYSAYVTASTRFGDGKTSNIIISFOTPEG-PSOPKDV 956  
QY 956 YYANLSSSIILFWTPPSKPGNGIIQYYSVYRNTSGTFMQNFTLHELINDFNMTVSTII 1015  
DB 957 YYANLSSSIILFWTPPSKPGNGIIQYYSVYRNTSGTFMQNFTLHEVNDNFDMVTSTII 1016  
QY 1016 DKLTIFSYTYFWLTASTSVGNKSSDIIEVYTDQDIPGFGVGNLTYESISSTAINVSW 1075  
DB 1017 DKLTIFSYTYFWLTASTSVGNKSSDIIEVYTDQDIPGFGVGNLTYESISSTAINVSW 1076  
QY 1076 PPAQNGLVFVYVSLILOQTPRHVRPPLVTVYERSIYFDNLEKYTDYILKITPSTKGFSD 1135  
DB 1077 PPAQNGLVFVYVSLILOQTPRHVRPPLVTVYERSIYFDNLEKYTDYILKITPSTKGFSD 1136  
QY 1136 TYTAQLYIKTEBDVDPETSPIINTFKNLSTSVLLSWDPPVPKNGAIISYDLTLQPNENY 1195  
DB 1137 TYTAQLYIKTEBDVDPETSPIINTFKNLSTSVLLSWDPPVPKNGAIISYDLTLQPNENY 1196  
QY 1196 SPITSDNYIILEELSPFTLYSFAAARTKGI GPSSILFEYTDSEVPPLAPPONLTLINCT 1255  
DB 1197 SPITSDNYIILEELSPFTLYSFAAARTKGI GPSSILFEYTDSEVPPLAPPONLTLINCT 1256  
QY 1256 SDFVWLKWSPLPGGIVKVYSFKIHEHETDTIYKNI SGFKTEAKLVGLEPVSTYSIRV 1315  
DB 1257 SDFVWLKWSPLPGGIVKVYSFKIHEHETDTIYKNI SGFKTEAKLVGLEPVSTYSIRV 1316  
QY 1316 SAFTKVGNGNOPSNNVKTTOESVDPVQVQNCMAATSWQSVLVKWDPPKANGIITQYMW 1375  
DB 1317 SAFTKVGNGNOPSNNVKTTOESVDPVQVQNCMAATSWQSVLVKWDPPKANGIITQYMW 1376  
QY 1376 TVERNSTKVSPODHMYTFIKLLANTSYVFKVRAASTSAGEBDESTCHVSTLPSTVSPVPTN 1435  
DB 1377 TVERNSTKVSPODHMYTFIKLLANTSYVFKVRAASTSAGEBDESTCHVSTLPSTVSPVPTN 1436  
QY 1436 IAFSDVQSTSATLTWIRPDTILGYFQNYKITTLQRAQCKEWESECEVEYQKIQLYBAH 1495  
DB 1437 IAFSDVQSTSATLTWIRPDTILGYFQNYKITTLQRAQCKEWESECEVEYQKIQLYBAH 1496  
QY 1496 LTEETVYGLKPRWTRPOVAASTNAGYGNASNWISTKTLPGPDPPPPNVHVVAATSPFSI 1555  
DB 1497 LTEETVYGLKPRWTRPOVAASTNAGYGNASNWISTKTLPGPDPPPPNVHVVAATSPFSI 1556  
QY 1556 SISWSEPAVITGPTCYLIDVKSVDNDEFNIFIKSNEENKTIIEIKDLEIFTRYSVVITAF 1615  
DB 1557 SISWSEPAVITGPTCYLIDVKSVDNDEFNIFIKSNEENKTIIEIKDLEIFTRYSVVITAF 1616  
QY 1616 TGNISAAVEGKSSAEMIVTTLESAPDPPNNMTFKIPDEVTFKQLTFLPSPQNGNIQ 1675  
DB 1617 TGNISAAVEGKSSAEMIVTTLESAPDPPNNMTFKIPDEVTFKQLTFLPSPQNGNIQ 1676  
QY 1676 VYQALVYREDDPTAVQIHNLSIIQKNTFTVIAMLEGLKGHTYNTISVAVNSAGAPKVP 1735  
DB 1677 VYQALVYREDDPTAVQIHNLSIIQKNTFTVIAMLEGLKGHTYNTISVAVNSAGAPKVP 1736  
QY 1736 MRLTMDIKAPARPKTKPTPIYDATCKLAVTSTTIIRMPICVYSDHDPKIKNVQVLATET 1795  
DB 1737 MRLTMDIKAPARPKTKPTPIYDATCKLAVTSTTIIRMPICVYSDHDPKIKNVQVLATET 1796  
QY 1796 GAQHDGNTYKWDAYFNKARPFTNEGFPNPPCTEGKTKFSGNEBIYIIGADNACMIPGN 1855  
DB 1797 GAQHDGNTYKWDAYFNKARPFTNEGFPNPPCTEGKTKFSGNEBIYIIGADNACMIPGN 1856  
QY 1856 EDKICNGPLKPKQVLFKFRATNIMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCL 1915  
DB 1857 EDKICNGPLKPKQVLFKFRATNIMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCL 1916



```
QY 1916 SIILLGTAIPAFARIKOKEGTYSPOAEIITDKLKDQITVADLEKDLRLTR--- 1972
Db 1917 SIILLGTAIPAFARIKOKEGTYSPOAEIITDKLKDQITVADLEKDLRLTRLS 1976
QY 1973 -----PISKKSFLQVBEELCTNNNLKFOBEFSELPKFLQDLSSSTDADLPWRAKORFPN 2026
Db 1977 YRSIKPISKKSFLQVBEELCTNNNLKFOBEFSELPKFLQDLSSSTDADLPWRAKORFPN 2036
QY 2027 IKPYNNNRVKLIADASVPGSDYINASYISGVLCPNEFIATQGPLGTGDFWVMWETR 2086
Db 2037 IKPYNNNRVKLIADASVPGSDYINASYISGVLCPNEFIATQGPLGTGDFWVMWETR 2095
QY 2087 AKTLVMTQCFEGRIRCHOYWPEDNKPVTVGDIVITKLMEDVQIDWTIRDLKIERHGD 2146
Db 2096 AKTLVMTQCFEGRIRCHOYWPEDNKPVTVGDIVITKLMEDVQIDWTIRDLKIERHGD 2155
QY 2147 CMTVROCNFTANPEHGVSPENSAPLIHFVKLVASRAHDTTPMLVHCSAGVGRGTGVFIAD 2206
Db 2156 CMTVROCNFTANPEHGVSPENSAPLIHFVKLVASRAHDTTPMLVHCSAGVGRGTGVFIAD 2215
QY 2207 HLTOHNDHDFVDIYGLVAELRSERCMVONLAQYIFLHCILDLNLSNKGSNQPICFVNY 2266
Db 2216 HLTOHNDHDFVDIYGLVAELRSERCMVONLAQYIFLHCILDLNLSNKGSNQPICFVNY 2275
QY 2267 SALQKMDSLDAMEGDVELEWEETM 2291
Db 2276 SALQKMDSLDAMEGDVELEWEETM 2300
```

## RESULT 5

```
US-10-466-759-2
; Sequence 2, Application US/10466759
; Publication No. US20040081983A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: LEE, Ernestine A.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: ISON, Craig H.
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: YAO, Monique G.
; APPLICANT: JACKSON, Jennifer L.
; APPLICANT: TANG, Y. Tom
; APPLICANT: YUE, Henry
; APPLICANT: TRAN, Bao
; APPLICANT: DING, Li
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: LAL, Preeti G.
; TITLE OF INVENTION: KINASES AND PHOSPHATASES
; FILE REFERENCE: PI-0344 USN
; CURRENT APPLICATION NUMBER: US/10/466,759
; CURRENT FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: PCT/US02/01369
; PRIOR FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: US 60/263,083
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 60/271,205
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/271,117
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/276,859
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/278,504
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/278,522
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/280,510
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: US 60/280,266
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
```

```
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 2299
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7480588CD1
US-10-466-759-2

Query Match      98.4%; Score 11797; DB 15; Length 2299;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 2272; Conservative 6; Mismatches 5; Indels 24; Gaps 8;

QY 1 MDPLIIFLLPFGTSETQVDVSNVPGTRDYDITISISITSTYTPVTRIIVTPNWKPGPPV 60
Db 1 MDPLIIFLLPFGTSETQVDVSNVPGTRDYDITISISITSTYTPVTRIIVTPNWKPGPPV 55
QY 61 FLAGERVGSAGILLSWNTPPNNGRIISYIVKYKVCPMQVTVTVQVRSKPSDLEVLN 120
Db 56 FLAGERVGSAGILLSWNTPPNNGRIISYIVKYKVCPMQVTVTVQVRSKPSDLEVLN 115
QY 121 LNPGETTYEIKVAEENSAGIGVSDPFLFQTAESAAGKVVNLTVAEYNASAVKLIWLP 180
Db 116 LNPGETTYEIKVAEENSAGIGVSDPFLFQTAESAAGKVVNLTVAEYNASAVKLIWLP 175
QY 181 PNGKITSPKISVKHARSGIWKDVIRVEDILTGLPECNENSESFLWSTASPSPTLGRV 240
Db 176 PNGKITSPKISVKHARSGIWKDVIRVEDILTGLPECNENSESFLWSTASPSPTLGRV 235
QY 241 TPPSRTHSSSTLTONEISSWKEPISFVWTHLRPYTTLPEVSAATTEAGYIDSTIVRT 300
Db 236 TPPSRTHSSSTLTONEISSWKEPISFVWTHLRPYTTLPEVSAATTEAGYIDSTIVRT 295
QY 301 PESVPEGPONCVTCNITGKSPSILLWDPPTIVTGFSYRVLYGSPS--GRILDNSTKDLK 358
Db 296 PESVPEGPONCVTCNITGKSPSILLWDPPTIVTGFSYRVLYGSPSAGRILDNSTKDLK 355
QY 359 FAFTNLTPPTWYDVYIAAETSAGTGPKGNISVFTPPDVGAVDFDLQAEVSTQVRAITWK 418
Db 356 FAFTNLTPPTWYDVYIAAETSAGTGPKGNISVFTPPDVGAVDFDLQAEVSTQVRAITWK 415
QY 419 KRPQNGIINQYRVKVLVPETGIIILENTLLTGNNEYINDPMAPEIYVIVPVMVLYEGSA 478
Db 416 KRPQNGIINQYRVKVLVPETGIIILENTLLTGNNE-INDPMAPEIYVIVPVMVLYEGSA 474
QY 479 EMSSDLHSLATPIYNSHPDKNPPARNRAEDQTSPPVVTTRNOYITDIAAQSLSVIRRLVP 538
Db 475 EMSSDLHSLATPIYNSHPDKNPPARNRAEDQTSPPVVTTRNOYITDIAAQSLSVIRRLVP 534
QY 539 FTEHMIYSAPFTIMGEGPPTVLSVTRQVPSIKIINYKNISSSSILLYWDPPEYPNGK 598
Db 535 FTEHMIYSAPFTIMGEGPPTVLSVTRQVPSIKIINYKNISSSSILLYWDPPEYPNGK 594
QY 599 ITHYTIYAMELDTNRAFOITTDNSFLITG----LKKYTKYKRVAAASTHDSSELSSEN 654
Db 595 ITHYTIYAMELDTNRAFOITTDNSFLITGGRKWLKKYTKYKRVAAASTHDSSELSSEN 654
QY 655 DIFVRTSDEPSSQDQVEIDVTADEIRLKWSPKPNGLIIAYEVLKNIDTLTKMKT 714
Db 655 DIFVRTSDEPSSQDQVEIDVTADEIRLKWSPKPNGLIIAYEVLKNIDTLTKMKT 714
QY 715 STTDIILNLRPHLTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDAPENITYKNISSG 774
Db 715 STTDIILNLRPHLTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDAPENITYKNISSG 774
QY 775 EIELSFLPPSSPNGIIKKYTIYVKASNGNEERTINTTSLTONIKVKKYQVYIEVSAST 834
Db 775 EIELSFLPPSSPNGIIKKYTIYVKASNGNEERTINTTSLTONIKVKKYQVYIEVSAST 834
QY 835 LKGEGRSAPISILTEEDAPDPPQDPSVKQLSGVTVKLSWQPPLEPNGLIILYTYVVM- 893
Db 835 LKGEGRSAPISILTEEDAPDPPQDPSVKQLSGVTVKLSWQPPLEPNGLIILYTYVVMR 894
```

```
QY 894 NRSSKTIINVTETSLDLDYNEVSAYVTASTREFGDKTGTSNIIISFOTPEGASDPK 953
Db 895 NRSSKTIINVTETSLDLDYNEVSAYVTASTREFGDKTRSNIIISFOTPEG-PSDPK 953
QY 954 DVYANLSSSSIIILFWTPSPKNGIIQYISVYRYNTSGTFMQNFTLHETLNDPDMVTST 1013
Db 954 DVYANLSSSSIIILFWTPSPKNGIIQYISVYRYNTSGTFMQNFTLHEVTNDPDMVTST 1013
QY 1014 IIDKLTIFSYTFLWTASTSVGNKSSDIIIEVYTDQDIPGFGVGNLTYESISSSTAINVS 1073
Db 1014 IIDKLTIFSYTFLWTASTSVGNKSSDIIIEVYTDQDVPFGVGNLTYESISSSTAINVS 1073
QY 1074 WPPAQPNGLVFYVSLIIQOTPRHVRPLVTVYESIYFDNLEKTYDYLKLTTPSTKGP 1133
Db 1074 WPPAQPNGLVFYVSLIIQOTPRHVRPLVTVYESIYFDNLEKTYDYLKLTTPSTKGP 1133
QY 1134 SDTYTAQLYIKTEEDVPETSPIINTPKNLSSVLLSWDPVPKNGAIIISYDLTLQGPNE 1193
Db 1134 SDTYTAQLYIKTEEDVPETSPIINTPKNLSSVLLSWDPVPKNGAIIISYDLTLQGPNE 1193
QY 1194 NYSFITSNDYIIIEELSPPTLSYFFAAARTKGLGPSSILFFYTDDESVPPLAPPQNLTLN 1253
Db 1194 NYSFITSNDYIIIEELSPPTLSYFFAAARTKGLGPSSILFFYTDDESVPPLAPPQNLTLN 1253
QY 1254 CTSDFVWLKWSPLPGGIVKYVSPKIHETDITIIYKNI SGFKTEAKLVGLEPVSYSI 1313
Db 1254 CTSDFVWLKWSPLPGGIVKYVSPKIHETDITIIYKNI SGFKTEAKLVGLEPVSYSI 1313
QY 1314 RVSFAFTKVGNGQFNVVAFKTTQESVPDVQNNQCMATSWQSVLVKWDPPKKANGIIQY 1373
Db 1314 RVSFAFTKVGNGQFNVVAFKTTQESVPDVQNNQCMATSWQSVLVKWDPPKKANGIIQY 1373
QY 1374 MVTVERNSTKVPQDHYTFIKLLANTS VFKVRASTSAGEDESTCHVSTLPTETVPSVP 1433
Db 1374 MVTVERNSTKVPQDHYTFIKLLANTS VFKVRASTSAGEDESTCHVSTLPTETVPSVP 1433
QY 1434 TNAFSDVOSTNTLWIPDITLGFQNYKITTQRAOKKEWSEECVEYKQIYLVE 1493
Db 1434 TNAFSDVOSTNTLWIPDITLGFQNYKITTQRAOKKEWSEECVEYKQIYLVE 1493
QY 1494 AHLTEBTVGLKFRWYRQVAASTNAGYCNASNISTKTLPQPPGPPENHVWATSPF 1553
Db 1494 AHLTEBTVGLKFRWYRQVAASTNAGYCNASNISTKTLPQPPGPPENHVWATSPF 1553
QY 1554 SISISWSEPAVITGPTCYLIDVKSVDNDFNISFKISNENKTIETIKOLEIFTRYSVWIT 1613
Db 1554 SISISWSEPAVITGPTCYLIDVKSVDNDFNISFKISNENKTIETIKOLEIFTRYSVWIT 1613
QY 1614 AFTGNISAAVYEGKSAEMIVTTLESAPKDPNNMTFQKIPDEVTKFQLTFLPPSPNGN 1673
Db 1614 AFTGNISAAVYEGKSAEMIVTTLESAPKDPNNMTFQKIPDEVTKFQLTFLPPSPNGN 1673
QY 1674 IOVYQALVYREDDPTAVQIHNLSIIQKTNFTVIALEGLKGHTYNI SVYAVNSAGAGPK 1733
Db 1674 IOVYQALVYREDDPTAVQIHNLSIIQKTNFTVIALEGLKGHTYNI SVYAVNSAGAGPK 1733
QY 1734 VPMRITMDIKAPARKTKPTPIYDATGKLLVSTTTITIRMPICYSSDDHGPIKNVOVLAT 1793
Db 1734 VPMRITMDIKAPARKTKPTPIYDATGKLLVSTTTITIRMPICYSSDDHGPIKNVOVLAT 1793
QY 1794 ETGAQHDGNTWKYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEIIYIIGADNACMIP 1853
Db 1794 ETGAQHDGNTWKYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEIIYIIGADNACMIP 1853
QY 1854 GNEDKICNGPLPKPKQYLFKFRATNIMGQFTDSYDDPVKTLGEGLSERTVEIILSVTLC 1913
Db 1854 GNEDKICNGPLPKPKQYLFKFRATNIMGQFTDSYDDPVKTLGEGLSERTVEIILSVTLC 1913
QY 1914 ILSIILLGTAIFAFARIQKQEGGYSPQDAEIIDTKLKLQDLITVADLELKDRLTR- 1972
Db 1914 ILSIILLGTAIFAFARIQKQEGGYSPQDAEIIDTKLKLQDLITVADLELKDRLTR- 1973

QY 1973 -----PISKSKFLQHVHEELCTNNNLKFOBEFSELPKFLQDLSSTADLPWNAKQNF 2024
Db 1974 LSYRSKSIKPIKSKSFLQHVHEELCTNNNLKFOBEFSELPKFLQDLSSTADLPWNAKQNF 2033
QY 2025 PNIKPYNNNRVVKLIADASVPGSDYINASYISGYLCNPEFIATQGPLPGTVGDFWRMWE 2084
Db 2034 PNKEY-NNNRVKLIADASVPGSDYINASYISGYLCNPEFIATQGPLPGTVGDFWRMWE 2092
QY 2085 TRAKTLVMTQCFEKGRIKCHQWPEDNKPVTVFGDIVITKLMEVDQIDWTIRDLKIERH 2144
Db 2093 TRAKTLVMTQCFEKGRIKCHQWPEDNKPVTVFGDIVITKLMEVDQIDWTIRDLKIERH 2152
QY 2145 GDCMTRVRCNFTAWPEHGVPENSAPLIHFVKLVRASRAHDTTPMIVHCSAGVGRGVFTA 2204
Db 2153 GDCMTRVRCNFTAWPEHGVPENSAPLIHFVKLVRASRAHDTTPMIVHCSAGVGRGVFTA 2212
QY 2205 LDHLTOHINDHDFVDIYGLVAELRSERCMQVNTAQYIFLHQICILDLLSNKGSNQPICFV 2264
Db 2213 LDHLTOHINDHDFVDIYGLVAELRSERCMQVNTAQYIFLHQICILDLLSNKGSNQPICFV 2272
QY 2265 NYSALQKQSDSLDAMEGDVELEWEETM 2291
Db 2273 NYSALQKQSDSLDAMEGDVELEWEETM 2299

RESULT 6
US-10-087-684-6
; Sequence 6, Application US/10087684
; Publication No. US20040029116A1
; GENERAL INFORMATION:
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: MacDougall, John R.
; APPLICANT: Millet, Isabelle
; APPLICANT: Ellerman, Karen
; APPLICANT: Stone, David J.
; APPLICANT: Grosse, William M.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgesse, Catherine E.
; APPLICANT: Casman, Stacie, J.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Li, Li
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Mishra, Vishnu
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Rastelli, Luca
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Miller, Charles E.
; APPLICANT: Gangolli, Esha A.
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-214 CIP
; CURRENT APPLICATION NUMBER: US/10/087,684
; CURRENT FILING DATE: 2003-03-10
; PRIOR APPLICATION NUMBER: 60/253,834
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/250,926
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/264,180
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,656
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/327,456
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 6
```

; LENGTH: 2281  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-087-684-6

Query Match 93.4%; Score 11191; DB 15; Length 2281;  
Best Local Similarity 94.5%; Pred. No. 0;  
Matches 217; Conservative 40; Mismatches 63; Indels 24; Gaps 16;

QY 1 MDLFIPLLLFIGTSETQVDSNVVPGTRVDITISSITTTTSPVTRIVTPNTKPGPPV 60  
DB 1 MDLFIPLLLFIGTSETQVDSNVVPGTRVDITISSITTTTSPVTRIGASN--EPGPPV 58  
QY 61 FLAGERVGSAGILLSNWTPNPNKGRILSIYVKVEVCPNMOTVTVQVRSKPDGLEVLTN 120  
DB 59 FLAGERVGSAGILLSNWTPNPNKGRILSIYVKVEVCPNMOTVTVQVRSKPDGLEVLTN 118  
QY 121 LNPCTTVEIKVAEASAGIVFSDPFLFOFAESAPGVKNLTVAEYNASAVKLIWLPQ 180  
DB 119 LNPCTTVEIKVAEASAGIVFSDPFLFOFAESAPGVKNLTVAEYNASAVKLIWLPQ 177  
QY 181 PNGKITSPKISVKHARSIGVVKDVSIRVEDILTKLP-ECNENSESFLMSTASPSPTLGR 239  
DB 178 KK-KITSFKISVKHNRSGIVVKEVSIRVECILSASLPHCNENSESFLMSTASPSPTLGR 236  
QY 240 VTPPSRTHSSSTLTONEISSVWKBPISFVVVTHLRPVTYTLFVVSAAATGAYIDSTIVR 299  
DB 237 VTPPSRTHSSSTLTONEISSV-KBPISFVVVTHLRPVTYTLFVVSAAATGAYIDSTIVR 295  
QY 300 TPESVPEGPONCVTGNITKFSIILWDPPTIVTGKFSYRVELYGPS-GRILDNSTKDLK 358  
DB 296 TPESVPEGPONCVTGNITKFSIILWDPPTIVTGKFSYRVELYGPSAGRILDNSTKDLK 355  
QY 359 FATNLTPPTMYDVYIAAETSAGTGPKSNISVTPDPVCAVFDLQALAEVSTQVRI TWK 418  
DB 356 FATNLTPPTMYDVYIAAETSAGTGPKSNISVTPDPVCAVFDLQALAEVSTQVRI TWK 415  
QY 419 KRPQNGIINOYRKVKLVPTGTIILENTLLTGNNVINDMPAPEI NVIPEPMVGLYEGSA 478  
DB 416 KRPQNGIINOYRKVKLVPTGTIILENTLLTGNN-INDMPAPEI NVIPEPMVGLYEGSA 474  
QY 479 EMSDDLHLATFYNSHPDKNFARNRAEDQTSFVVTTRNOYITDIAEQLSVVIRLVP 538  
DB 475 EMSDDLHLATFYNSHPDKNFARNRAEDQTSFVVTTRNOYITDIAEQLSVVIRLRR 534  
QY 539 FTEHMLSVSAFTIMGEGPPTVLSVTRQOVPSIKIINLYK--NISSSILLYWDPRPEYN 596  
DB 535 FWAETGFSRYTITMSS-----ASRNLTSFGPLSAQNFAVTHVITTEVFLHWDPPD--P 586  
QY 597 GKITHYTIYAMELDT-NRAFOITIDN-SFLITGLKKYTKYKRVAAASTHDGESSISEEN 654  
DB 587 VPFHHVLTILDVENQSKSILATLNSLSLVGLKKYTKYKRVAAASTHVGESSISEEN 646  
QY 655 DIFVRSSEDPSSPOQVEIDVTADIRLKWSPPEKNGIIIAEYVLKNDITLYMKNT 714  
DB 647 DIFVRSSEDPSSPOQVEIDVTADIRLKWSPPEKNGIIIAEYVLKNDITLYMKNT 706  
QY 715 STTDIILRNLRPHTLYNISVRSYTRFCHGNQVSSLSVRSSETPVDSAPENITYKNISG 774  
DB 707 STTDIILRNLRPHTLYNISVRSYTRFCHGNQVSSLSVRSSETPVDSAPENITYKNISG 766  
QY 775 EIELSFLPPSSNGIIKKYTYILKRSNGNEERTINTTSLTONIKVLKKTQYIIIEVSAST 834  
DB 767 EIELSFLPPSSNGIIKKYTYILKRSNGNEERTINTTSLTONIKVLKKTQYIIIEVSAST 826  
QY 835 LKGEGRVSAPI SILTEDAPSPQDFSVKQLSGVTVKLSQWPLPENGIIILYTYVWN 894  
DB 827 LKGEGRVSAPI SILTEDAPSPQDFSVKQLSGVTVKLSQWPLPENGIIILYTYVWN- 885  
QY 895 RSSLKTINVTETSLSLDLDYNYEYSAVYTASTRFGDKTGSNIISFOTPEGAPSPPKD 954  
DB 886 RSSLKTINVTETSLSLDLDYNYEYSAVYTASTRFGDKTGSNIISFOTPEG-PSDPPKD 944

QY 955 VYYANLSSSIILFWTPPSKPNGIIQYYSVYVRNTSGTFMONTLHEL TNDPDMTVSTI 1014  
DB 945 VYYANLSSSIILFWTPPSKPNGIIQYYSVYVRNTSGTFMONTLHEL TNDPDMTVSTI 1004  
QY 1015 IDKLITIFSYYTFLWTASTSVGNKSSDIEVYTDQDIEGFGVGNLTYSISSTAINVSW 1074  
DB 1005 IDKLITIFSYYTFLWTASTSVGNKSSDIEVYTDQDIEGFGVGNLTYSISSTAINVSW 1064  
QY 1075 VPPAQNGLVFYVYVLSIILQOTPRHVRPPLVTVERSYIFDNLEKYTDYIILKITESTEGFS 1134  
DB 1065 VPPAQNGLVFYVYVLSIILQOTPRHVRPPLVTVERSYIFDNLEKYTDYIILKITESTEGFS 1124  
QY 1135 DYTAAQYIKTEEDVPETSPINTPKNLSTSVLLSWDPVVKNGAIISVDLTLOQPNEN 1194  
DB 1125 DYTAAQYIKTEEDVPETSPINTPKNLSTSVLLSWDPVVKNGAIISVDLTLOQPNEN 1184  
QY 1195 YSPITSDNVIILBELSPFTLYSFFFAAARTRKGLGPSIILFFYTDSEVPLAPPQNLTINC 1254  
DB 1185 YSPITSDNVIILBELSPFTLYSFFFAAARTRKGLGPSIILFFYTDSEVPLAPPQNLTINC 1244  
QY 1255 TSDFVWLKWSPPGLPGGIVKVYSFKIHEHETDTIYKNI SGFKTEAKLVGLEPVSTYSIR 1314  
DB 1245 TSDFVWLKWSPPGLPGGIVKVYSFKIHEHETDTIYKNI SGFKTEAKLVGLEPVSTYSIR 1304  
QY 1315 VSAFTKVGNGNFSNVVKTTOESVDPVQVQNMCMATSWQSULVKWDPKPKANGIITQYM 1374  
DB 1305 VSAFTKVGNGNFSNVVKTTOESVDPVQVQNMCMATSWQSULVKWDPKPKANGIITQYM 1364  
QY 1375 VTVERNSTKVSQDDHMYTFIKLLANTSYVFKVRASTASAGEDSTCHVSTLPETVSVPT 1434  
DB 1365 VTVERNSTKVSQDDHMYTFIKLLANTSYVFKVRASTASAGEDSTCHVSTLPETVSVPT 1424  
QY 1435 NIAFSDVQSTATLWIRPDTILGYFQNYKIITQLRAQCKWESECEVEYQIKQYLYEA 1494  
DB 1425 NIAFSDVQSTATLWIRPDTILGYFQNYKIITQLRAQCKWESECEVEYQIKQYLYEA 1484  
QY 1495 HLTEETVYGLKFRWTRFOVAASTNAGYNASNWISTKTLPGPDGPPENHVAVTSPPS 1554  
DB 1485 HLTEETVYGLKFRWTRFOVAASTNAGYNASNWISTKTLPGPDGPPENHVAVTSPPS 1544  
QY 1555 ISISSEPAVITGPTCYLIDVKSVDNDEFNISPISKEENKTIRIKDLIELFTRYSVVITA 1614  
DB 1545 ISISSEPAVITGPTCYLIDVKSVDNDEFNISPISKEENKTIRIKDLIELFTRYSVVITA 1604  
QY 1615 FTGNTSAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKQLTFLPPSQNGNI 1674  
DB 1605 FTGNTSAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKQLTFLPPSQNGNI 1664  
QY 1675 QYVQALVYREDDPTAVQIHNLISIIQKNTTFVIAMLEGLKGGHTYINSVYAVNSAGAPKV 1734  
DB 1665 QYVQALVYREDDPTAVQIHNLISIIQKNTTFVIAMLEGLKGGHTYINSVYAVNSAGAPKV 1724  
QY 1735 PMRITMDIKAPARKTKPTPIYDATGKLLVTSTTITIRMPICIYSDHGPINKVQVLTAE 1794  
DB 1725 PMRITMDIKAPARKTKPTPIYDATGKLLVTSTTITIRMPICIYSDHGPINKVQVLTAE 1784  
QY 1795 TGAQHDGNTYKWDAYFNKARPYFTNEGPNPCTEGKTKFSGNEBIYIIGADNACMI PG 1854  
DB 1785 TGAQHDGNTYKWDAYFNKARPYFTNEGPNPCTEGKTKFSGNEBIYIIGADNACMI PG 1844  
QY 1855 NEDKI CNGPLKPKQYLFKFRATNIMGQFTSDSYSDPVKTLGEGLSERTVEIILSVTLCI 1914  
DB 1845 NEDKI CNGPLKPKQYLFKFRATNIMGQFTSDSYSDPVKTLGEGLSERTVEIILSVTLCI 1904  
QY 1915 LSIILLGTAFAPARIQKQEGGTYSPQDAI IITKLLDQITVADIELKDERLTRPI 1974  
DB 1905 LSIILLGTAFAPARIQKQEGGTYSPQDAI IITKLLDQITVADIELKDERLTRPI 1964  
QY 1975 SKKSFLOHVEELCTNNLKFQSEFSELPKFLQDLSTADLPMNRAKNFPNKPNNNN 2034  
DB 1965 SKKSFLOHVEELCTNNLKFQSEFSELPKFLQDLSTADLPMNRAKNFPNKP-NNN 2023  
QY 2035 RVKLIADASVPGSDVINASYISGYLCPNEFIATQGPLCTGVGDFWRMNVWETRAKTLVMLT 2094

Db 2024 RVKLIADASVPGSDYINASYISGYLCPNEFIATQGLPGTVGDFWVMVWETRAKTLVMLT 2083  
Qy 2095 QCQEKGRIRCHQWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVRQCN 2154  
Db 2084 QCQEKGRIRCHQWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVRQCN 2143  
Qy 2155 FTAPWPHGVGPNASAPLIHFVKLVASRAHDTPMIVHCSAGVGRGTGFALDHLTOHIND 2214  
Db 2144 FTAPWPHGVGPNASAPLIHFVKLVASRAHDTPMIVHCSAGVGRGTGFALDHLTOHIND 2203  
Qy 2215 HDFVDIYGLVAELSRMCMQVQLAQYIFLHQICLDLLSNKSNQPCFVNTSALQKXMS 2274  
Db 2204 HDFVDIYGLVAELSRMCMQVQLAQYIFLHQICLDLLSNKSNQPCFVNTSALQKXMS 2263  
Qy 2275 LDAME-GDVELEWEETM 2291  
Db 2264 LDAMEGGDVELEWEETM 2281

RESULT 7

US-10-218-779-6

Sequence 6, Application US/10218779  
Publication No. US20040029222A1

GENERAL INFORMATION:

APPLICANT: Edinger, Shlomit  
APPLICANT: MacDougall, John  
APPLICANT: Millet, Isabelle  
APPLICANT: Ellerman, Karen  
APPLICANT: Stone, David  
APPLICANT: Gerlach, Valerie  
APPLICANT: Grosse, William  
APPLICANT: Alsbrook II, John  
APPLICANT: Lepley, Denise  
APPLICANT: Rieger, Daniel  
APPLICANT: Burgess, Catherine  
APPLICANT: Casman, Stacie  
APPLICANT: Spytek, Kimberly  
APPLICANT: Boldog, Ferenc  
APPLICANT: Li, Li  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Mishra, Vishnu  
APPLICANT: Patturajan, Meera  
APPLICANT: Shenoy, Suresh  
APPLICANT: Rastelli, Luca  
APPLICANT: Tchernev, Velizar  
APPLICANT: Vernet, Corine  
APPLICANT: Zerhusen, Bryan  
APPLICANT: Malyankar, Uriel  
APPLICANT: Guo, Xiaojia  
APPLICANT: Miller, Charles  
APPLICANT: Gangolli, Esha

TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-214

CURRENT APPLICATION NUMBER: US/10/218,779

CURRENT FILING DATE: 2002-08-14

PRIOR APPLICATION NUMBER: 60/253,834

PRIOR FILING DATE: 2000-11-29

PRIOR APPLICATION NUMBER: 60/250,-926

PRIOR FILING DATE: 2000-11-30

PRIOR APPLICATION NUMBER: 60/264,180

PRIOR FILING DATE: 2001-01-25

PRIOR APPLICATION NUMBER: 60/313,656

PRIOR FILING DATE: 2001-08-20

PRIOR APPLICATION NUMBER: 60/327,456

PRIOR FILING DATE: 2001-10-05

NUMBER OF SEQ ID NOS: 216

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 6

LENGTH: 2281

TYPE: PRT

ORGANISM: Homo sapiens

US-10-218-779-6

Query Match 93.4%; Score 11191; DB 15; Length 2281;  
Best Local Similarity 94.5%; Pred. No. 0;  
Matches 2171; Conservative 40; Mismatches 63; Indels 24; Gaps 16;

Qy 1 MDPLIIFLLFLFTIGTSETQVDVSNVVPGRYDITISSISTYTSPTVTRIVTPNVRKPGPV 60  
Db 1 MDPLIIFLLFLFTIGTSETQVDVSNVVPGRYDITISSISTYTSPTVTRIGASN-ERGPVP 58  
Qy 61 FLAGERVSGAGILLSNWTPPNPNGRISIVIVKYKEVCPWMQTVYTVQVRSKPSLEVLITN 120  
Db 59 FLAGERVSGAGILLSNWTPPNPNGRISIVIVKYKEVCPWMQTVYTVQVRSKPSLEVLITN 118  
Qy 121 LNPQTTYRITKVAEASAGIGVSPDFLQTAESAPGKVVNLTVEAYNASAVKLIWVLPQ 180  
Db 119 LNPQTTYRITKVAEASAGIGVSPDFLQTAESAPGKVVDFTEAVPSS-KLMTYSAT 177  
Qy 181 PNGKITSPKISVKHARSGIVKDVSIKVEDILATGKLPECNENSESFLWSTASPSPTLGR 239  
Db 178 KK-KITSPKISVKHNRSGIVKVESIRVECILSASLPLHCNENSESFLWSTASPSPTLGR 236  
Qy 240 VTTPSRTHSSSTLTQNEISSVWKKEPISPFVWTHLRPYTYLFEVSAATTEAGYIISTIVR 299  
Db 237 VTTPSRTHSSSTLTQNEISSV-KEPISFVWTHLRPYTYLFEVSAATTEAGYIISTIVR 295  
Qy 300 TPESVPEGPPQNCVNTGKSFSLMDPPTIVTGKFSYRVLYGSPS-GRILDNSTKDLK 358  
Db 296 TPESVPEGPPQNCVNTGKSFSLMDPPTIVTGKFSYRVLYGSPSAGRIIDNSTKDLK 355  
Qy 359 FAFNTLTPFTMYDVYIAAETSAGTGPKSGNISVFTPPDVPAGVDFDLQLAEVESQVRIITWK 418  
Db 356 FAFNTLTPFTMYDVYIAAETSAGTGPKSGNISVFTPPDVPAGVDFDLQLAEVESQVRIITWK 415  
Qy 419 KPRQNGIINOYRVKVLVPETGIIILENTLLTGNEYINDPMAPEIVNIVPEWGLYEGSA 478  
Db 416 KPRQNGIINOYRVKVLVPETGIIILENTLLTGNEE-INDPMAPEIVNIVQPMVGLYEGSA 474  
Qy 479 EMSDDLHSLATFIYNSHPDKPPARNRAEDQTSPPVVTTRNOYITDIAEOLSVYIRLVR 538  
Db 475 EMSDDLHSLATFIYNSHPDKPPARNRAEDQTSPPVVTTRNQYITDIAEOLSVYIRLVR 534  
Qy 539 FTEHMSVSAFTIMGEGPPTVLSVTRTQVPSIKIINYK--NISSSILLWDPPEYN 596  
Db 535 FWAETMGFSRYTIMSS-----ASRDLNLTSPGLSAQNFRTVTHVITEVFLHWDPPD--P 586  
Qy 597 GKITHYTIYAMELDT-NRAFOITITDN-SFLITGLKKTYYKMYRVAASHTDESSISEEN 654  
Db 587 VFFHYLITILDVENQSKSIILRLNLSLSVLIGLKYYTKYKMRVAASHTHVGESSLSEEN 646  
Qy 655 DIFVTSDEPSSPDQVEVIDVTADERLKWSPPEKNGIILAYEVLKNTDLYMKNT 714  
Db 647 DIFVTSDEPSSPDQVEVIDVTADERLKWSPPEKNGIILAYEVLKNTDLYMKNT 706  
Qy 715 STTDIILNLRPHLTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSPAENITYKNISG 774  
Db 707 STTDIILNLRPHLTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSPAENITYKNISG 766  
Qy 775 EIELSFLPPSSPNGIILKYYTYLKRSGNEERTINTTSLTQNIKVLKKTQYIIIEVSAST 834  
Db 767 EIELSFLPPSSPNGIILKYYTYLKRSGNEERTINTTSLTQNIKVLKKTQYIIIEVSAST 826  
Qy 835 LKGEGRVSAPISILTEEDAPSPDQFVSKQSGVTVKLSWOPPLEPNCIILYYTVVYN 894  
Db 827 LKGEGRVSAPISILTEEDAPSPDQFVSKQSGVTVKLSWOPPLEPNCIILYYTVVYN 885  
Qy 895 RSSLATINVTETSLSDLDYNNVEYSAYVTASTREFGDKGTGNSNIIISFQTEGAPSDPPK 954  
Db 886 RSSLATINVTETSLSDLDYNNVEYSAYVTASTREFGDKGTGNSNIIISFQTEGAPSDPPK 944  
Qy 955 VYIANLSSSSIIILFWTPPSKNGIILQYYSVYTRNTSGTFMQNFTLHETLNDNDNTVSTI 1014  
Db 945 VYIANLSSSSIIILFWTPPSKNGIILQYYSVYTRNTSGTFMQNFTLHETLNDNDNTVSTI 1004

1015 IDKLTIFSYVYFWLTASTSVGNKSSDIIEVYTDQDIPEGFVGNLTYSISSTAINVSW 1074  
Db IDKLTIFSYVYFWLTASTSVGNKSSDIIEVYTDQDQPEGFVGNLTYSISSTAINVSW 1064  
1075 VPPAQNGLVFVYVSVLILOQTQPHRVPLVTVYERSIYFONLEKYDYLKIITPSTKGFPS 1134  
Db VPPAQNGLVFVYVSVLILOQTQPHRVPLVTVYERSIYFONLEKYDYLKIITPSTKGFPS 1124  
1135 DYTATQALYIKTEBDVPETSPINFTKVLSTSVLLSWDPVPKNGAIISYDLTLQGPNN 1194  
Db DYTATQALYIKTEBDVPETSPINFTKVLSTSVLLSWDPVPKNGAIISYDLTLQGPNN 1184  
1195 YSFITSDNVIILBELSPFTLYSFAAARTKGLGRSSILFFYTDQSVPLAPQNLTLINC 1254  
Db YSFITSDNVIILBELSPFTLYSFAAARTKGLGRSSILFFYTDQSVPLAPQNLTLINC 1244  
1255 TSDPFWLKWSPSPGPIVKVYFKIHEHETDIYVKNISGFKTEAKVLGLBPVSYISIR 1314  
Db TSDPFWLKWSPSPGPIVKVYFKIHEHETDIYVKNISGFKTEAKVLGLBPVSYISIR 1304  
1315 VSAFTKVGNGQPSNVVVKFTQESVDPVQNMCMATSWQSVLVKWDPPKANGIITQYM 1374  
Db VSAFTKVGNGQPSNVVVKFTQESVDPVQNMCMATSWQSVLVKWDPPKANGIITQYM 1364  
1375 VTVERNSTKVSQDQHMVYTIKLLANTSYVYKVRASAGGDESTCHVSTLPETVPSVPT 1434  
Db VTVERNSTKVSQDQHMVYTIKLLANTSYVYKVRASAGGDESTCHVSTLPETVPSVPT 1424  
1435 NIAPSDVQSTSATLWIRPDTILGYFQNYKIITQLRQAQCKEWESECEYOKIQLYEA 1494  
Db NIAPSDVQSTSATLWIRPDTILGYFQNYKIITQLRQAQCKEWESECEYOKIQLYEA 1484  
1495 HLTEETVYGLKFRWYRFOVAASTNAGYGNASNIWISTKLPDGPDPGPPENVHVAATSPS 1554  
Db HLTEETVYGLKFRWYRFOVAASTNAGYGNASNIWISTKLPDGPDPGPPENVHVAATSPS 1544  
1555 ISISWSEPAVITGPTCYLIDVKSVDNDEFNIFSIKNEENKTEIKDLBIFRYSVVITA 1614  
Db ISISWSEPAVITGPTCYLIDVKSVDNDEFNIFSIKNEENKTEIKDLBIFRYSVVITA 1604  
1615 FTGNTSAAVVEGKSSAEMIIVTLESAPKDPNNMTFQKIPDEVTKFQLTFLPSPQNGNI 1674  
Db FTGNTSAAVVEGKSSAEMIIVTLESAPKDPNNMTFQKIPDEVTKFQLTFLPSPQNGNI 1664  
1675 QYVQALVYREDDPTAQVHNLISIIOKTNTFVJAMLEGLKGHTYNI SVYAVNSAGAGPKV 1734  
Db QYVQALVYREDDPTAQVHNLISIIOKTNTFVJAMLEGLKGHTYNI SVYAVNSAGAGPKV 1724  
1735 PMRITMDIKAPARPKTPPIYDATGKLLVSTTITIRMPICYSDDHGPIKNVQVLAPE 1794  
Db PMRITMDIKAPARPKTPPIYDATGKLLVSTTITIRMPICYSDDHGPIKNVQVLAPE 1784  
1795 TGAQHDGNTWKYDAVFNKARPYFTNEGPNPCTEGKTKPSGNEBIYIIGADNACMIPG 1854  
Db TGAQHDGNTWKYDAVFNKARPYFTNEGPNPCTEGKTKPSGNEBIYIIGADNACMIPG 1844  
1855 NEDKICNGLPKPKQYLFKFRATNMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCI 1914  
Db NEDKICNGLPKPKQYLFKFRATNMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCI 1904  
1915 LSIILGTAIFAPARIQOKGEGTVSPQDABIIDTKLKLQDITVADI ELKDERLTRPI 1974  
Db LSIILGTAIFAPARIQOKGEGTVSPQDABIIDTKLKLQDITVADI ELKDERLTRPI 1964  
1975 SKKSPLOHVEELCTNNLKFQEFSEL PKFLODLSSTDADLPWNAKRPNNI KVNNNN 2034  
Db SKKSPLOHVEELCTNNLKFQEFSEL PKFLODLSSTDADLPWNAKRPNNI KVNNNN 2023  
2035 RVKLADASVPGSDYINASYISGYLCPNEFIATQGPLGVGD FWRVWVWETRAKTLVMLT 2094  
Db RVKLADASVPGSDYINASYISGYLCPNEFIATQGPLGVGD FWRVWVWETRAKTLVMLT 2083  
2095 QCFEKGRIRCHQWPEDNKPVTVFGDIVITKLMEDVQIDMTIRDLKIERHGDQMTVRQCN 2154

2084 QCFEKGRIRCHQWPEDNKPVTVFGDIVITKLMEDVQIDMTIRDLKIERHGDQMTVRQCN 2143  
QY 2155 FTAWPEHGVPENSAPLIHFVKLVRAASRAHDTTTPMIVHCSAGVGRTCVFIALDHLTQHIND 2214  
Db 2144 FTAWPEHGVPENSAPLIHFVKLVRAASRAHDTTTPMIVHCSAGVGRTCVFIALDHLTQHIND 2203  
QY 2215 HDPVDIYGLVABLRSERMCMQNLQAIYIFLHQICILDLNKGNSQBPICFVNYSALQKQMS 2274  
Db 2204 HDPVDIYGLVABLRSERMCMQNLQAIYIFLHQICILDLNKGNSQBPICFVNYSALQKQMS 2263  
QY 2275 LDAMB-GDVELEWEETTM 2291  
Db 2264 LDAMEGDDVELEWEETTM 2281

## RESULT 8

US-09-822-871-4  
; Sequence 4, Application US/09822871  
; Publication No. US20030099942A1  
; GENERAL INFORMATION:  
; APPLICANT: WEBSTER, Marion et al  
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CL001219  
; CURRENT APPLICATION NUMBER: US/09/822,871  
; CURRENT FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 2301  
; TYPE: PRP  
; ORGANISM: Rat  
US-09-822-871-4

Query Match 87.9%; Score 10536; DB 10; Length 2301;

Best Local Similarity 85.9%; Pred. No. 0;

Matches 1977; Conservative 162; Mismatches 151; Indels 12; Gaps 4;

QY 1 MDPLIIFLLFLLTGTSQVDSNVVPGTRYDITISSIS-TTYSPTVTRIVTNVTKPGPP 59

Db 1 MDHFHSFLLLTGTSQVDSNVVPGTRYDITISSIS-TTYSPTVTRIVTNVTKPGPP 60

QY 60 VFLAGERSGAGILLSWNTPPNPNNGRIISYIVKYEVCPPMOTVYTVQVSKPDSLEVLIT 119

Db 61 VFLAGERSGAGILLSWNTPPNPNNGRIISYIVKYEVCPPMOTVYTVQVSKPDSLEVLIT 120

QY 120 NLNPGTTTTEIKVAENAGIGVFSDFPLQTAESAPGVKVNLTVRAYNASAVKLIWYLR 179

Db 121 NLNPGTTTTEIKVAENAGIGVFSDFPLQTAESAPGVKVNLTVRAYNASAVKLIWYLR 180

QY 180 QPNKTIKTSKLVKHSARSGIVVKQYSIRVEDILTKGLPECNENSSFLWSTASPSPLGR 239

Db 181 QPNKTIKTSKLVKHSARSGIVVKQYSIRVEDILTKGLPECNENSSFLWSTASPSPLGR 240

QY 240 VTPPSRTTSSSTLQNEISSVWKEPISFVVTHTLPPYTYLFEVSAATTEAGYIDSTIVR 299

Db 241 VTPPSRTTSSSTLQNEISSVWKEPISFVVTHTLPPYTYLFEVSAATTEAGYIDSTIVR 300

QY 300 TPESVPEGPPQNCVNTGNITGKSFSLWDPPTIVTKFSYRVELYGPSGRILONSTKDLKP 359

Db 301 TPESVPEGPPQNCVNTGNITGKSFSLWDPPTIVTKFSYRVELYGPSGRILONSTKDLKP 360

QY 360 APTNLTPTMTVDVYTAETSAGTKPSKNSISVTPDPVGVADQLQAEVSTQVRLTWK 419

Db 361 APTNLTPTMTVDVYTAETSAGTKPSKNSISVTPDPVGVADQLQAEVSTQVRLTWK 420

QY 420 PRQNGIINQVRKVLVPEPTEGIIENLTLLTGNNEINDEMPAIEIVNIVPEPMVGLYEGSAB 479

Db 421 PRQNGIINQVRKVLVPEPTEGIIENLTLLTGNNEINDEMPAIEIVNIVPEPMVGLYEGSAB 480

QY 480 MSSDLHSLATFIYNSHPDKNFPARNABEDQTSFVVVTRNQYITDIAEQLSVVIRLVPF 539

|    |      |  |      |
|----|------|--|------|
|    |      | :       :     :   :   :   :   :   :   :   :   :   :   :   :   :  |      |
| Db | 481  | MSSDLHSPASFIYNSHPHNDPFASTRABEOSSPVVTRNQYMTDITAEQSLYSVVRLVPF      | 540  |
| Qy | 540  | TEHMIYSVAFTIMEGEPPTVLVSVRTROQVPSSIKNYKNISSSILLYWDPPEYPNGKI       | 599  |
| Db | 541  | TEHTISVAFTIMEGEPPTVLVTRTROQVPSSIQIINYKNISSSILLYWDPPEYPNGKI       | 600  |
| Qy | 600  | THYTIYAMELDTNRAFOITTIDNSFLITGLKKYTKYKORVAASTHDGESSLSSENDIFVR     | 659  |
| Db | 601  | THYTIYATELDTNRAFQMWTVDNSFLITGLKKYTRYKWRVAASTHVGESSLSENDIFVR      | 660  |
| Qy | 660  | TSEDEPPSSODVEIVDTADEIRLKWSPEKPNGIIIAIEVLYKNIIDTILYMKNTSTDI       | 719  |
| Db | 661  | TPEDEPPSSQDVQTVGSPELRKLWSPPEKPNGIIIAIEVLYQNADTLFVKNTSTDI         | 720  |
| Qy | 720  | ILANLRPHLTNYTSVRSVTFGHCGNVSSLLSVRTSETVPDSPAPENITYKNISSGRELS      | 779  |
| Db | 721  | IISDLKEPYLIYNISIRSYTRLGHGNQSLLSVRTSETVPDSPAPENITYKNISSGRELS      | 780  |
| Qy | 780  | FLPPSSPNGIIKKYTIYLKESNGNEERTINTTSLTONIKVLKKYTOYIIEVSASTLKGE      | 839  |
| Db | 781  | FLPPRSPNGIIQKYTIYLKESNSHEARTINTTSLTOTIGLKKYTHVIEVSASTLKGE        | 840  |
| Qy | 840  | VRGAPISILTEEDAPSDPOFSVKQLSGVTYVKLSWPQPLEPNGIILYYTYVWNRSCLK       | 899  |
| Db | 841  | IRSRPISILTEEDAPSDPONFVSVKQLSGVTYVMLS WPQPLEPNGIILYYTYVMDKSSLR    | 900  |
| Qy | 900  | TINVTETSELSDLDYNNVESAYVTASTRGDGKTGNSIIISFOTPEGAPSDPKOVVYAN       | 959  |
| Db | 901  | AINATEASLVSLUDJYNDYGACVATRTRGDGNARSSIIINFRTPEGESDPDNDVHYVN       | 960  |
| Qy | 960  | LSSSSIILFWTPPSKPNGIIOYTSVYYRNTSGTFMQNFTHELTNDDNMVTSTIIDKLT       | 1019 |
| Db | 961  | LSSSSIILFWTPPKPNGIIOYTSVYYONTSGTFVQNFLLQVTKESDNVTSARIYLA         | 1020 |
| Qy | 1020 | IFSYYTWLTASTS VGVNGNKSSDIIIEVYTQDIPEGFVGMLTYEISSSTAINVSWVPPAQ    | 1079 |
| Db | 1021 | IFSYYTWLTASTS VGVNGNKSSDIIHYVTQDIPEGFVGMLTFESISSTAIHSVWEPPSQ     | 1080 |
| Qy | 1080 | PNGLVFYVSLILOOT-PRVURPLVATYBESIYFDNLKETYDYTLKLTPTSEKGFSPTYT      | 1138 |
| Db | 1081 | PNGLVFYLSLNQQSPRHMIPLVYENSIDPDLEKETYDYFKLTPTSEKGFSPTYT           | 1140 |
| Qy | 1139 | AQYIKTEEDVPETSPIINTFKNLSSTSVLLSWDPVPVKPGNAIIISDYLTIQGPNENYSFI    | 1198 |
| Db | 1141 | TQLHIKTEEDVPDTPPIINTFKNLSSILLSWDPLKPGAULGYHLTLQGPANHFTV          | 1200 |
| Qy | 1199 | TSOYIIILEBSPPTLSYFFAAATRKG LGPSSILFFYTDESVP LAPQMLTINCTSDF       | 1258 |
| Db | 1201 | TSGNHIVLEBSPPTLSYFFAAATMKG LGPSSILFFYTDESAP LAPQMLTINYTSDF       | 1260 |
| Qy | 1259 | VWLKWSPLPGGITVKYYSKIHHEHTDTIYKNI SGFKTEAKLVGLEPSTYSIRVASF        | 1318 |
| Db | 1261 | VWLTWSPSPFGGITVKYYSFKIHHEHTDTVFYKNI SGLQDAKLEGLEPSTYSVSASF       | 1320 |
| Qy | 1319 | TKVGNHQFNSNVKFTTQESVPDVVONQC MATSQSVLVKWDPKKANGIIIOYMYTVE        | 1378 |
| Db | 1321 | TKVGNHQYSNVBEFTTQESVPEAVRNIECVARDWQSVSVRWDPDKRTNGII IHYMITVG     | 1380 |
| Qy | 1379 | RNSTKVSPDHMYTIFIKLLANTSYVFKVRASTSAGEDESTCHVSTLPETVPSPVTNI AF     | 1438 |
| Db | 1381 | GNS TKVSPRDPTYTFTKLLPNTSYVFEVRASTSAGENE SRCDOI STL PETVPSPATNVAF | 1440 |
| Qy | 1439 | SDVOGSTATLTWIPDPDILGFONYKITTLQRAQCKEWESERCEVOKI OYLYEAHLTE       | 1498 |
| Db | 1441 | SNVOGSTATLTWTPKDPTIFGYFONYKITTLQRAQCKREWEPRECEIHKQDOYLYEANQTE    | 1500 |
| Qy | 1499 | ETVYGLKKFRWYRFQVAASTNAGYGNASNMI STKLP GPDPGPPENVHVATSPFSISIS     | 1558 |
| Db | 1501 | ETVHGLKKFRWYRFQVAASTNVGYSNASEMI STQTL GPDPGPPENVHVATSPFGINIS     | 1560 |
| Qy | 1559 | WSEPAVITGPTCYLIDVKSVDNDBFNIFSPIKSNENKTI EBIKOLEIFTRYSVVITAFTGN   | 1618 |

|      |     |        |       |       |       |      |      |       |     |     |     |     |     |     |    |     |      |    |    |    |    |     |     |     |    |      |      |    |    |      |      |   |   |   |   |   |   |   |    |    |   |    |      |   |    |      |   |      |   |   |   |   |      |      |   |      |      |
|------|-----|--------|-------|-------|-------|------|------|-------|-----|-----|-----|-----|-----|-----|----|-----|------|----|----|----|----|-----|-----|-----|----|------|------|----|----|------|------|---|---|---|---|---|---|---|----|----|---|----|------|---|----|------|---|------|---|---|---|---|------|------|---|------|------|
| 1561 | WS  | PAVITG | TFYIL | LDVKS | VDDDD | FNIS | FLKS | NEENK | TT  | INN | LEV | TP  | YS  | VIT | PA | VGN | 1621 |    |    |    |    |     |     |     |    |      |      |    |    |      |      |   |   |   |   |   |   |   |    |    |   |    |      |   |    |      |   |      |   |   |   |   |      |      |   |      |      |
| 1619 | ISA | Y      | VEG   | KS    | AE    | MI   | VTT  | LES   | AP  | KD  | PNN | NMT | FQ  | I   | P  | DEV | TK   | FO | LT | PL | PS | QPN | GNI | QV  | YQ | 1678 |      |    |    |      |      |   |   |   |   |   |   |   |    |    |   |    |      |   |    |      |   |      |   |   |   |   |      |      |   |      |      |
| 1621 | VS  | RAY    | TG    | K     | SS    | AE   | VI   | IT    | LES | VP  | KD  | PNN | NMT | FQ  | I  | P   | DEV  | TK | FO | LT | PL | PS  | QPN | GNI | RV | YQ   | 1680 |    |    |      |      |   |   |   |   |   |   |   |    |    |   |    |      |   |    |      |   |      |   |   |   |   |      |      |   |      |      |
| 1679 | AL  | VY     | RED   | DP    | TA    | VO   | I    | HN    | S   | II  | OK  | NT  | FN  | VI  | AM | LE  | G    | KG | HT | Y  | NI | S   | VI  | AV  | NS | AG   | AK   | VP | MR | 1738 |      |   |   |   |   |   |   |   |    |    |   |    |      |   |    |      |   |      |   |   |   |   |      |      |   |      |      |
| 1681 | AL  | VY     | RED   | DP    | TA    | VO   | I    | HN    | S   | II  | OK  | NT  | FN  | VI  | AM | LE  | G    | KG | HT | Y  | NI | S   | VI  | AV  | NS | AG   | AK   | VP | Q  | MR   | 1740 |   |   |   |   |   |   |   |    |    |   |    |      |   |    |      |   |      |   |   |   |   |      |      |   |      |      |
| 1739 | TMD | I      | K     | A     | P     | AK   | P    | K     | T   | P   | I   | Y   | D   | AT  | G  | K   | L    | L  | T  | S  | T  | T   | I   | R   | M  | P    | C    | Y  | S  | D    | D    | H | G | P | I | K | N | V | OL | AT | E | GA | 1798 |   |    |      |   |      |   |   |   |   |      |      |   |      |      |
| 1741 | TMD | I      | K     | A     | P     | AK   | P    | K     | S   | K   | P   | I   | P   | I   | R  | D   | AT   | G  | K  | L  | L  | T   | S   | T   | T  | I    | R    | M  | P  | C    | Y    | N | D | D | H | G | P | I | R  | N  | V | OL | AT   | E | GA | 1800 |   |      |   |   |   |   |      |      |   |      |      |
| 1799 | H   | D      | G     | N     | V     | T    | K    | W     | Y   | D   | A   | Y   | F   | N   | K  | A   | P    | Y  | T  | N  | E  | G   | F   | N   | P  | P    | C    | T  | E  | G    | K    | T | F | S | G | N | E | E | Y  | I  | I | G  | A    | D | N  | A    | C | M    | I | P | G | N | E    | D    | K | 1858 |      |
| 1801 | Q   | D      | G     | N     | V     | T    | K    | W     | Y   | D   | A   | Y   | F   | N   | K  | A   | P    | Y  | T  | N  | E  | G   | F   | N   | P  | P    | C    | I  | E  | G    | K    | T | F | S | G | N | E | E | Y  | I  | I | G  | A    | D | N  | A    | C | M    | I | P | G | N | E    | E    | K | 1860 |      |
| 1859 | I   | C      | N     | G     | L     | P    | K    | P     | K   | Q   | Y   | L   | F   | K   | F  | R   | A    | T  | N  | I  | M  | G   | O   | F   | T  | S    | D    | S  | D  | P    | V    | K | T | L | E | G | L | S | R  | T  | V | E  | I    | L | S  | V    | T | L    | C | I | L | S | I    | 1918 |   |      |      |
| 1861 | I   | C      | N     | G     | L     | P    | K    | P     | K   | Q   | Y   | L   | F   | K   | F  | R   | A    | T  | N  | I  | M  | G   | O   | F   | T  | S    | D    | S  | D  | P    | V    | K | T | L | E | G | L | S | R  | T  | V | E  | I    | L | S  | V    | T | L    | C | I | L | S | I    | 1920 |   |      |      |
| 1919 | L   | L      | G     | T     | A     | I    | F    | A     | P   | A   | R   | I   | R   | O   | K  | E   | G    | T  | Y  | S  | P  | O   | D   | A   | E  | I    | D    | T  | K  | L    | D    | O | L | I | T | V | A | D | L  | E  | K | D  | B    | E | L  | R    | T | 1972 |   |   |   |   |      |      |   |      |      |
| 1921 | L   | L      | G     | T     | A     | I    | F    | A     | P   | A   | R   | I   | R   | O   | K  | E   | G    | T  | Y  | S  | P  | R   | D   | A   | E  | I    | D    | T  | K  | L    | D    | O | L | I | T | V | A | D | L  | E  | K | D  | B    | E | L  | R    | T | L    | S | Y | R | K | 1980 |      |   |      |      |
| 1973 | -   | -      | -     | P     | I     | S    | K    | S     | F   | L   | O   | H   | V   | E   | L  | C   | T    | N  | N  | N  | L  | K   | F   | O   | E  | F    | S    | E  | L  | P    | K    | F | O | D | L | S | T | D | A  | D  | I | P  | N    | N | A  | K    | R | F    | N | I | K | P | 2029 |      |   |      |      |
| 1981 | S   | I      | R     | P     | I     | S    | K    | S     | F   | L   | O   | H   | V   | E   | L  | C   | T    | N  | N  | N  | L  | K   | F   | O   | E  | F    | S    | E  | L  | P    | K    | F | O | D | L | S | T | D | A  | D  | I | P  | N    | N | A  | K    | R | F    | N | I | K | P | 2040 |      |   |      |      |
| 2030 | Y   | N      | N     | N     | R     | V    | K    | L     | I   | A   | D   | A   | S   | V   | P  | G   | S    | D  | I  | N  | A  | S   | I   | S   | G  | Y    | L    | C  | P  | N    | E    | F | I | A | T | O | Q | P | L  | E  | G | T  | V    | G | D  | F    | R | W    | M | Y | E | T | R    | A    | K | T    | 2089 |
| 2041 | Y   | -      | N     | N     | N     | R    | V    | K     | L   | I   | A   | D   | V   | S   | L  | P   | G    | S  | D  | I  | N  | A   | S   | I   |    |      |      |    |    |      |      |   |   |   |   |   |   |   |    |    |   |    |      |   |    |      |   |      |   |   |   |   |      |      |   |      |      |

```

RESULT 9
US-10-673-885-4
; Sequence 4, Application US/10673885
; Publication No. US20040081644A1
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001219DIV
; CURRENT APPLICATION NUMBER: US/10/673,885
; CURRENT FILING DATE: 2003-09-30
; PRIOR APPLICATION NUMBER: 09/822,871
; PRIOR FILING DATE: 04-02-2001
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2301
; TYPE: PRT
; ORGANISM: Rat
US-10-673-885-4

```

Query Match 87.9%; Score 10536; DB 15; Length 2301;  
Best Local Similarity 85.9%; Pred. No. 0;  
Matches 1977; Conservative 162; Mismatches 151; Indels 12; Gaps 4;

QY 1 MFLIIFLLFICTSETQDVSNVPTGTYDITISS-ITTSPTVRIWNTVTKGPP 59  
DB 1 MDHPSFLFLICTSESQDVSSPDGTGYDITLSSVATTYSSPVSRILTATVTKGPP 60

QY 60 VFLAGRVSGAGILLSWNTPPNNGRIISYVYKKEVCPWQVYTVQVRSKPSLVLIT 119  
DB 61 VFLAGRVSGAGILLSWNTPPNNGRIISYVYKKEVCPWQVYTVQVRSKPSLVLIT 120

QY 120 NLMPTTYEIKVAENSAGIGVSDPFLPOTASAPGKVVNLTVAYNASAVKLWYLR 179  
DB 121 NLMPTTYEIKVAENNAGIGVSDPFLPOTASAPGKVVNLTVAYNASAVKLWYLR 180

QY 180 QPNKITSFKISVKHARSIGVVDVSRVEDIITLGLPECNENSESFLMSTASPSFLGR 239  
DB 181 QPNKITSFKISVKHARSIGVVDVSRVEDIITLGLPECNENSESFLMSTASPSFLGR 240

QY 240 VTPPSRTTHSSPLTQNEISSVWKEPISFVTHLRPYTYLPEVSAATAGYIDSTIVR 299  
DB 241 VTPPSRTTHSSPLTQNEISSVWKEPISFVTHLRPYTYLPEVSAATAGYIDSTIVR 300

QY 300 TPESVPEGPPQNCVTGNITCKSPESILWDPTIIVTGKFSYRVLYGSGRILDNSTKDLF 359  
DB 301 TPESVPEGPPQNCVTGNITCKSPESILWDPTIIVTGKFSYRVLYGSGRILDNSTKDLF 360

QY 360 APTNLPTFYDVIYIAAETSAGTGPKNISVFTPPDVGAFDLOLAEVESQVRIWTK 419  
DB 361 APTNLPTFYDVIYIAAETSAGTGPKNISVFTPPDVGAFDLOLAEVESQVRIWTK 420

QY 420 PRQNGIINQYRVKVLVPEGTGIIILENTLLTGNNEYINDPMAPEIVNIBPVMGLYEGSAE 479  
DB 421 PRQNGIINQYRVKVLVPEGTGIIILENTLLTGNNEYINDPMAPEIVNIBPVMGLYEGSAE 480

QY 480 MSSDLHSLATFYNHPDKNFPARNRAEDOTSVPVTRNQYITDIAAOLSVIRRLVFP 539  
DB 481 MSSDLHSLATFYNHPDKNFPARNRAEDOTSVPVTRNQYITDIAAOLSVIRRLVFP 540

QY 540 TEHMIVSAPTIMGEPPTVLSVTRQOVPSSTIKIINYNISSSSILLYWDPEYNGKI 599  
DB 541 TEHMIVSAPTIMGEPPTVLSVTRQOVPSSTIKIINYNISSSSILLYWDPEYNGKI 600

QY 600 THYTIYAMELDTNRAQITTDINSFLITGLAKYTKYKMRVAASHTDGSLSSENDIFVR 659  
DB 601 THYTIYAMELDTNRAQITTDINSFLITGLAKYTKYKMRVAASHTDGSLSSENDIFVR 660

QY 660 TSEDEPSSQDVVEIDVTADEIRLKWSPPEKNGIIIAEVLVYKNIIDLYMKNTSTDI 719  
DB 661 TSEDEPSSQDVVEIDVTADEIRLKWSPPEKNGIIIAEVLVYKNIIDLYMKNTSTDI 720

QY 720 ILRLNLPHTLYNISVRSYTRFGHGNQVSSLSVRTSETVPDSAPENITYKNISSGEIELS 779  
DB 721 ILRLNLPHTLYNISVRSYTRFGHGNQVSSLSVRTSETVPDSAPENITYKNISSGEIELS 780

QY 780 FLPPSPNGIIXKYTYILKRSNGNEBRTINTSLTONIKVLKYTOYIIIEVSASTLKGG 839  
DB 781 FLPPSPNGIIXKYTYILKRSNGNEBRTINTSLTONIKVLKYTOYIIIEVSASTLKGG 840

QY 840 VRSAPISILTEADAPSPQDFSVKOLSGVTVKLSWQPLEPENGIIILYTVVYNNRSSIX 899  
DB 841 VRSAPISILTEADAPSPQDFSVKOLSGVTVKLSWQPLEPENGIIILYTVVYNNRSSIX 900

QY 900 TINVETSLESLDLDYNNVEYSAVTASTRGDKTGSNIIISQTPGAPSDPPKOVYAN 959  
DB 901 TINVETSLESLDLDYNNVEYSAVTASTRGDKTGSNIIISQTPGAPSDPPKOVYAN 960

QY 960 LSSSSIIILFWTPPSKENGIIQYYSVYRYNTSGTFMONEFTLHETDNDNMTSTIIDKLT 1019  
DB 961 LSSSSIIILFWTPPSKENGIIQYYSVYRYNTSGTFMONEFTLHETDNDNMTSTIIDKLT 1020

QY 1020 IFSYTFWLTASTSVGNKSSDIIIEVYTDQDIPEGFVGNLTYESISSTAINVSWVPPAQ 1079

DB 1021 IFSYTFWLTASTSVGNKSSDIIIEVYTDQDIPEGFVGNLTYESISSTAINVSWVPPAQ 1080

QY 1080 PNLGVFYVYSLILQOT-PRHVRPPLVYTERSYFONLEKYTYILKITEPSTKGFSDTYT 1138  
DB 1081 PNLGVFYVYSLILQOT-PRHVRPPLVYTERSYFONLEKYTYILKITEPSTKGFSDTYT 1140

QY 1139 AQIYKTEEDVETSPINFPKLGASTSVLLSWDPKPKNGALISVDLTLQGNENYSP 1198  
DB 1141 TQLHITKTEEDVETSPINFPKLGASTSVLLSWDPKPKNGALISVDLTLQGNENYSP 1200

QY 1199 TSDNYIILBELSPETLYSPFAAARTKGLGPSSILFFYTDSEVPLAPPNLTINCTSD 1258  
DB 1201 TSDNYIILBELSPETLYSPFAAARTKGLGPSSILFFYTDSEVPLAPPNLTINCTSD 1260

QY 1259 VMLKSPSPPLGGIVKVSFKIHEHETDIYIKNISGFKTEAKLVGLBVSIVSYRVSF 1318  
DB 1261 VMLKSPSPPLGGIVKVSFKIHEHETDIYIKNISGFKTEAKLVGLBVSIVSYRVSF 1320

QY 1319 TKVGNQPSNVVKTQESVDPVQNMCMATSWQSVLVKWDPPKANGIITQYVMTVE 1378  
DB 1321 TKVGNQPSNVVKTQESVDPVQNMCMATSWQSVLVKWDPPKANGIITQYVMTVE 1380

QY 1379 RNSTKVSPODHMYTFIKLLANTSYVFKVRASTAGSDESTCHVSTLPTVSPVNTIAF 1438  
DB 1381 RNSTKVSPODHMYTFIKLLANTSYVFKVRASTAGSDESTCHVSTLPTVSPVNTIAF 1440

QY 1439 SDVQSTSATLWIRPDTILGYFQNYKITTQLRAQCKEWESECEVEYQIKIYLAHLTE 1498  
DB 1441 SDVQSTSATLWIRPDTILGYFQNYKITTQLRAQCKEWESECEVEYQIKIYLAHLTE 1500

QY 1499 ETYVGLKPRWTRFOVAASTNAGYGNASNIWISTKTLPGPDGPPENHVHVSPPSIS 1558  
DB 1501 ETYVGLKPRWTRFOVAASTNAGYGNASNIWISTKTLPGPDGPPENHVHVSPPSIS 1560

QY 1559 WSEPAVITGPTCYLVLDVKSVDNDENISFKIKNENKTIIEIKDLEIFTSYVITAFGN 1618  
DB 1561 WSEPAVITGPTCYLVLDVKSVDNDENISFKIKNENKTIIEIKDLEIFTSYVITAFGN 1620

QY 1619 ISAAVVEGKSSAEMTIVTLESAPKDPNNMTFKIPDEVTQFQTLFPLPSQPNNGINQVQ 1678  
DB 1621 ISAAVVEGKSSAEMTIVTLESAPKDPNNMTFKIPDEVTQFQTLFPLPSQPNNGINQVQ 1680

QY 1679 ALVYREDPTAVQIHNLSIIQKNTFVIAMLBGLKGGHTYNSVYAVNSAGAPKVPMRI 1738  
DB 1681 ALVYREDPTAVQIHNLSIIQKNTFVIAMLBGLKGGHTYNSVYAVNSAGAPKVPMRI 1740

QY 1739 TWDIIPAPKPKTPPIVDATGKLLVSTTIIIRMPICYSDDHGPIKVVQVLAETGAQ 1798  
DB 1741 TWDIIPAPKPKTPPIVDATGKLLVSTTIIIRMPICYSDDHGPIKVVQVLAETGAQ 1800

QY 1799 HDGNTWKYDAVFNKARPYFTNEGPNPCTSGKTKFSGNBEIYIIGADNACMI PGNEBK 1858  
DB 1801 HDGNTWKYDAVFNKARPYFTNEGPNPCTSGKTKFSGNBEIYIIGADNACMI PGNEBK 1860

QY 1859 ICNGLPKPKQYLFKFRATNIMGQTDSDYSPVTKLGEGLSERTVEIILSTLCILSI 1918  
DB 1861 ICNGLPKPKQYLFKFRATNIMGQTDSDYSPVTKLGEGLSERTVEIILSTLCILSI 1920

QY 1919 LIGTAIPAPARIQKQKGGTYSPODAEIIDTKLQDLITVADLELKDRLTR----- 1972  
DB 1921 LIGTAIPAPARIQKQKGGTYSPODAEIIDTKLQDLITVADLELKDRLTR----- 1980

QY 1973 ---PISSKSFLOHVEBELCTNNLKFQEFSELPKFLQDLSSDADLPNNRKNRPNNIKP 2029  
DB 1981 ---PISSKSFLOHVEBELCTNNLKFQEFSELPKFLQDLSSDADLPNNRKNRPNNIKP 2040

QY 2030 YNNNRVXKLADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFRWMWETRAKT 2089  
DB 2041 YNNNRVXKLADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFRWMWETRAKT 2099

QY 2090 LVMLTQCFEKGIRCHQYWPEDNKPVTVFGDITVITKLMEDVQIDMTIRDLKTIERHGDQMT 2149



Db 2100 LVMLTQCFKGRIRCHQYWPEDNKPVTVFGDIVITKLMEDIOIDWITRDLKTIERHGDGMT 2159  
QY 2150 VRQCNFTAMPPEHGVNPENSAPLHFVKLVRAASRAHDTTPMIVHCSAGVGRGTGVFIADHLT 2209  
Db 2160 VRQCNFTGPHGVNPENTPFLHFVKLVTSRAHDTTPMVHCSAGVGRGTGVFIADHLT 2219  
QY 2210 QHNDHDFVDIYGLVAELSRMCMVQNLAQVIFLHQICLDLLSNKGSNQPCFVNYNSAL 2269  
Db 2220 QHNDHDFVDIYGLVAELSRMCMVQNLAQVIFLHQICLDLLSNKGSNQPCFVNYNSTL 2279  
QY 2270 QKMSLDNAMEDGVLEWEETTM 2291  
Db 2280 QKMSLDNAMEDGVLEWEETTM 2301

RESULT 10  
US-10-087-684-37  
; Sequence 37, Application US/10087684  
; Publication No. US20040029116A1  
; GENERAL INFORMATION:  
; APPLICANT: Edinger, Shlomit R.  
; APPLICANT: MacDougall, John R.  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Stone, David J.  
; APPLICANT: Grose, William M.  
; APPLICANT: Lepley, Denise M.  
; APPLICANT: Rieger, Daniel K.  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Casman, Stacie, J.  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Boldog, Ferenc L.  
; APPLICANT: Li, Li  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Mishra, Vishnu  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Zerhusen, Bryan D.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Miller, Charles E.  
; APPLICANT: Gangolli, Basha A.  
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
; FILE REFERENCE: 21402-214 CIP  
; CURRENT APPLICATION NUMBER: US/10/087,684  
; PRIOR FILING DATE: 2003-03-10  
; PRIOR APPLICATION NUMBER: 60/253,834  
; PRIOR FILING DATE: 2000-11-29  
; PRIOR APPLICATION NUMBER: 60/250,926  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: 60/264,180  
; PRIOR FILING DATE: 2001-01-25  
; PRIOR APPLICATION NUMBER: 60/274,194  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/313,656  
; PRIOR FILING DATE: 2001-08-20  
; PRIOR APPLICATION NUMBER: 60/327,456  
; PRIOR FILING DATE: 2001-10-05  
; NUMBER OF SEQ ID NOS: 220  
; SOFTWARE: CuraSeqList version 0.1  
; SEQ ID NO 37  
; LENGTH: 2302  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-10-087-684-37

Query Match 87.9%; Score 10536; DB 15; Length 2302;  
Best Local Similarity 85.9%; Pred. No. 0;  
Matches 1977; Conservative 162; Mismatches 151; Indels 12; Gaps 4;  
QY 1 MDPLIFLLFICTSETQVDNSVNPVPGTRYDITISSIS-TTYSPTRIVTPNVTKPGRP 59

Db 2 MDPHFSFLLLIGTSESDVSSPDGTGYDITLLSSVSAITYSSFSVSTLATNVTKPGPP 61  
QY 60 VFLAGERVGSAGILLSWNTPNPNCRIISYIVKYKEVCPWMQTVTVQVRSKPDSEVLILT 119  
Db 62 VFLAGERVGSAGILLSWNTPNPNCRIISYIVKYKEVCPWMQTVATYTRAKRADSLEVLIT 121  
QY 120 NLNPCTTYEIKVAEENSAGIVFSDPFLFQTAESAPGKVNLTVEAYNASAVKLIIWYLP 179  
Db 122 NLNPCTTYEIKVAEENSAGIVFSDPFLFQTAESAPGKVNLTVEALNYSVNLIIWYLP 181  
QY 180 QNGKITSFKISVKHARSIGVIVKOVSVIRVEDILTOKLPECNENSSPFWSTASPSPTLGR 239  
Db 182 QNGKITSFKISVKHARSIGVIVKOVSVIRVEDILTOKLPECNENSSPFWSTTSPSTLGR 241  
QY 240 VTPPSRTTHSSSTLTQNEISSVWKEPISFVVTHTLRPYTYTLFVESAATTEAGYIDSTTVR 299  
Db 242 VTPVTRTTQSSSTAARSKISSVWKEPISFVVTHTLRPYTYTLFVESAATTEAGYIDSTTVR 301  
QY 300 TPESVPEGPQNCVTGNITGKSFILMDPPPIVTGKFSYRVELYGFSGRILDNSTKDLKF 359  
Db 302 TPESVPEGPQNCVTGNITGKSFILMDPPPIVTGKFSYRVELYGFSGRILDNSTKDLRF 361  
QY 360 APTNLTPPTMDVYIAETSAGTGPKSNISVFTPPDPVCGAVFDLQLAEVSTQVTRTWK 419  
Db 362 APTNLTPPTMDVYIAETSAGTGPKSNISVFTPPDPVCGAVFDLQLAEVSTQVTRTWK 421  
QY 420 PRQNGIINOYRVKVLVPETGILGENTLLTGNEVINDPMAPEIYVIVPEVMVGLVGSAB 479  
Db 422 PRQNGIINOYRVKVLVPETGILGENTLLTGNEVINDPMAPEIYVIVPEVMVGLVGSAB 481  
QY 480 MSSDLHSLATFIYNSHPDKNFPARNABDQSPVVVTRNQYITDIAAQLSIVIRRLVPP 539  
Db 482 MSSDLHSLATFIYNSHPDKNFPARNABDQSPVVVTRNQYITDIAAQLSIVIRRLVPP 541  
QY 540 TEHMTISVSAFTIMGEGPPTVLSVRTRQVQPSIKIINYNKISSSILLYWDPPEYNGKI 599  
Db 542 TEHMTISVSAFTIMGEGPPTVLSVRTRQVQPSIKIINYNKISSSILLYWDPPEYNGKI 601  
QY 600 THYTIYAMELDTNRAFOITTDNSFLITGLKKYTKYKRVAASTHGDGSSISEENDIFVR 659  
Db 602 THYTIYAMELDTNRAFOITTDNSFLITGLKKYTKYKRVAASTHGDGSSISEENDIFVR 661  
QY 660 TSEDEPSSPODEVIDVTADIEIRLKWSPPEKPGNIIIAEVLKKNITQYIIIEVSASTLKGE 719  
Db 662 TSEDEPSSPODEVIDVTADIEIRLKWSPPEKPGNIIIAEVLKKNITQYIIIEVSASTLKGE 721  
QY 720 ILRLNRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSGEIBLS 779  
Db 722 ILRLNRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSGEIBLS 781  
QY 780 FLPPSSPNGIILKYYIYLRKSNNGNEERTINTSLTONIKVLKKYTOYIIIEVSASTLKGE 839  
Db 782 FLPPSSPNGIILKYYIYLRKSNNGNEERTINTSLTONIKVLKKYTOYIIIEVSASTLKGE 841  
QY 840 VRSAPISILTEEDAPDPSPQVSVKQLSGVTVKLSWQPPLEPNGLIYTVVVMNRSSILK 899  
Db 842 VRSAPISILTEEDAPDPSPQVSVKQLSGVTVKLSWQPPLEPNGLIYTVVVMNRSSILK 901  
QY 900 TINTVETSLSELDLDYVVEYSAVYTAFTFRFGDKTGSNIIISFQTPGEGAPSDPKDYYVAN 959  
Db 902 TINTVETSLSELDLDYVVEYSAVYTAFTFRFGDKTGSNIIISFQTPGEGAPSDPKDYYVAN 961  
QY 960 LSSSSIIILFWTPPSKPGNIIQYISVYVYNTSGTFMONTLHETLNDFNMTVSTIIDKLT 1019  
Db 962 LSSSSIIILFWTPPSKPGNIIQYISVYVYNTSGTFMONTLHETLNDFNMTVSTIIDKLT 1021  
QY 1020 IFSYTYFWLTASTSVGNKSSDIIIEVYTDODIPEGFVGNLTYESISTAINVSVWPPAQ 1079  
Db 1022 IFSYTYFWLTASTSVGNKSSDIIIEVYTDODIPEGFVGNLTYESISTAINVSVWPPAQ 1081  
QY 1080 PNGLYFYVYSILQOT-PRHVRPPIVYTYERSIYFDNLEKYTDYILKITPSTKSGSDYTT 1138

Db 1082 PNLGVFYLSLNLQQSPRRHMIPLVYENSIDFDLLEKTYDVIKLTPTSTKGFSETVT 1141  
Qy 1139 AQYIKTEEDVPETSPINTPKNSLSTSVLLSWDPVPKPNKALISYDLTLQGNENYSFI 1198  
Db 1142 TQLHIKTEEDVPETSPINTPKNSLSTSVLLSWDPVPKPNKALISYDLTLQGNENYSFI 1201  
Qy 1199 TSDNYILIELSPPTLYSPFAARTKGLGPSSILPFTYDESPLAPQNLTLINCTSDP 1258  
Db 1202 TSGNHIVLELSPPTLYSPFAARTKGLGPSSILPFTYDESPLAPQNLTLINCTSDP 1261  
Qy 1259 VMLKWSPLPGGIVKVSFKIHEHETDTIYKNSIGFKTEAKLVLEPVSYSIRVSAP 1318  
Db 1262 VMLTWSPLPGGIVKVSFKIHEHETDTIYKNSIGFKTEAKLVLEPVSYSIRVSAP 1321  
Qy 1319 TKVGNQNSNVKFTQSVDPVQVQMCATSWOSVLVKMDPPKANGIITQYMYTVE 1378  
Db 1322 TKVGNQNSNVKFTQSVDPVQVQMCATSWOSVLVKMDPPKANGIITQYMYTVE 1381  
Qy 1379 RSTKVSQDHYMTFILLANTSYVFKVRASTAGSGDESTCHVSTLPETVSPVNIAP 1438  
Db 1382 GNSTKVSQDHYMTFILLANTSYVFKVRASTAGSGDESTCHVSTLPETVSPVNIAP 1441  
Qy 1439 SDVQSTSATLWTRPDTILGYFONYKITTLQRAQCKEWESECEVQKIYLYEAHLTE 1498  
Db 1442 SNVQSTSATLWTRPDTILGYFONYKITTLQRAQCKEWESECEVQKIYLYEAHLTE 1501  
Qy 1499 ETYVGLKFRWYRFOVAASNAGYGNASWISFTLPGPPDGPENNVHVATSPFSIS 1558  
Db 1502 ETYVGLKFRWYRFOVAASNAGYGNASWISFTLPGPPDGPENNVHVATSPFSIS 1561  
Qy 1559 WSEPAVITGPTCYLIDVKSVDNDEFNISFKSNEENKTIKIDLEIFTRYSVVITAFGN 1618  
Db 1562 WSEPAVITGPTCYLIDVKSVDNDEFNISFKSNEENKTIKIDLEIFTRYSVVITAFGN 1621  
Qy 1619 ISAYVEGKSSAMVITLESAPKDPNNMTFOKIDEVTKQLTLPSPQNGNITVQ 1678  
Db 1622 VSRAYTDGSSABVIIITLESVPKDPNNMTFOKIDEVTKQLTLPSPQNGNITVQ 1681  
Qy 1679 ALVYREDDTAVQIHLNLSIIQKNTFVIAMLEGLKGHTYNSVAVNSAGAGPKVPMRI 1738  
Db 1682 ALVYREDDTAVQIHLNLSIIQKNTFVIAMLEGLKGHTYNSVAVNSAGAGPKVPMRI 1741  
Qy 1739 TMDIKAPARKPTPIYDATGKLLTSTTITIRMPICYSDHGPKNVQVLATGTGAQ 1798  
Db 1742 TMDIKAPARKPTPIYDATGKLLTSTTITIRMPICYSDHGPKNVQVLATGTGAQ 1801  
Qy 1799 HDGNVTKWDAYFNKARPFTNEGFNPPCTEGTKFSGNEHYIYGADNACMIPGNEK 1858  
Db 1802 QDGNVTKWDAYFNKARPFTNEGFNPPCTEGTKFSGNEHYIYGADNACMIPGNEK 1861  
Qy 1859 ICGNPLKPKKQYLKFRATNIMGQFTSDSYSDPVKTLGGLSERTVEIILSVTLCLISII 1918  
Db 1862 ICGNPLKPKKQYLKFRATNIMGQFTSDSYSDPVKTLGGLSERTVEIILSVTLCLISII 1921  
Qy 1919 LLGTAIFAFARIRQKQEGTYSPODAEIIDTKLKLQDITVADLELKDRLTR 1972  
Db 1922 LLGTAIFAFARIRQKQEGTYSPODAEIIDTKLKLQDITVADLELKDRLTR 1981  
Qy 1973 ----PISKKSFLQHEVELCTNNLKFQEFSELPKFLQDLSSTDAIPLWNRKRPNTKIP 2029  
Db 1982 SIKPISKKSFLQHEVELCTNNLKFQEFSELPKFLQDLSSTDAIPLWNRKRPNTKIP 2041  
Qy 2030 YNNNRVKLIADASVPGSDYINASYISGVLCEFIATQGLPGTGVDFRWWMETRATK 2089  
Db 2042 Y-NNNRVKLIADASVPGSDYINASYISGVLCEFIATQGLPGTGVDFRWWMETRATK 2100  
Qy 2090 LVMLTQCFEKGIRCHQYWPENKPVTVFGDIVITKLMEDVQIDWTIRDLKERHGDCT 2149  
Db 2101 LVMLTQCFEKGIRCHQYWPENKPVTVFGDIVITKLMEDVQIDWTIRDLKERHGDCT 2160  
Qy 2150 VRQCNFTAWPERGVSPENSAPLIHFVKLVRASTRAHDTTPMIHCSAGVGRGTGVFIADHUT 2209  
Db 2161 VRQCNFTAWPERGVSPENSAPLIHFVKLVRASTRAHDTTPMIHCSAGVGRGTGVFIADHUT 2220

Qy 2210 QHINHDFVDIYGLVAELRSEKCMQVNLQAQYIFLHQICILLLSNKGSNQPICFVNSAL 2269  
Db 2221 QHINHDFVDIYGLVAELRSEKCMQVNLQAQYIFLHQICILLLSNKGSNQPICFVNSAL 2280  
Qy 2270 QKMSDILDAMEGDEVELEWETTM 2291  
Db 2281 QKMSDILDAMEGDEVELEWETTM 2302

## RESULT 11

US-10-218-779-37  
; Sequence 37, Application US/10218779  
; Publication No. US2004002922A1  
; GENERAL INFORMATION:  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: MacDougall, John  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Stone, David  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Grosse, William  
; APPLICANT: Alsobrook II, John  
; APPLICANT: Lepley, Denise  
; APPLICANT: Rieger, Daniel  
; APPLICANT: Burgess, Catherine  
; APPLICANT: Casman, Stacie  
; APPLICANT: Spytke, Kimberly  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Li, Li  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Mishra, Vishnu  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Tchernev, Velizar  
; APPLICANT: Vernet, Corine  
; APPLICANT: Zerhusen, Bryan  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Miller, Charles  
; APPLICANT: Gangolli, Esha  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-214  
; CURRENT APPLICATION NUMBER: US/10/218,779  
; CURRENT FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: 60/253,834  
; PRIOR FILING DATE: 2000-11-29  
; PRIOR APPLICATION NUMBER: 60/250,-926  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: 60/264,180  
; PRIOR FILING DATE: 2001-01-25  
; PRIOR APPLICATION NUMBER: 60/313,656  
; PRIOR FILING DATE: 2001-08-20  
; PRIOR APPLICATION NUMBER: 60/327,456  
; PRIOR FILING DATE: 2001-10-05  
; NUMBER OF SEQ ID NOS: 216  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 37  
; LENGTH: 2302  
; TYPE: PRT  
; ORGANISM: Rattus rattus

US-10-218-779-37  
Query Match 87.9%; Score 10536; DB 15; Length 2302;  
Best Local Similarity 85.9%; Pred. No. 0;  
Matches 1977; Conservative 162; Mismatches 151; Indels 12; Gaps 4;  
Qy 1 MDPLIIFLLIFIGSETQVDYSNVPGTRYDITISSIS-TTYSPTVTRIVTNVTKGPP 59  
Db 2 MDHFSLFLFLIGTSESQVDVSSPDGTYDITLSSVATTSVSPVSRATLATNVTKGPP 61  
Qy 60 VFLAGERSAGILLSWTTPNPNGRIISYIVKYKEVCPMMQTVTVQVRSPKDSLEVLIT 119

Db 62 VFLAGRVSGAGILLSWNTPPNPGRGIIISVVKYKEVCPWMTAVTRAKXPDLSLEVLIT 121  
QY 120 NLNPGTYTIEKVAENSAGIGVSPDPLFOTAEASBGKVVNLTVRAYNASAVKLIWYLP 179  
Db 122 NLNPGTYTIEKVAENNAGIGVSPDPLFOTAEASBGKVVNLTVREALNYSAVNLWYLP 181  
QY 180 QNGKITTSFKISVGHARSIGVVKDVSIRVEDILITGKLPECNENSBFLWSTASPSPTLGR 239  
Db 182 QNGKITTSFKISVGHARSIGVVKDVSIRVEDILITGKLPECNENSBFLWSTASPSPTLGR 241  
QY 240 VTPPSRTTHSSLTITONEISSVWKBPISFVWTHLAPYTYLFEVSAATTEAGYIDSTIVR 299  
Db 242 VTPVTRTQSSSTAARSKISSVWKBPISFVWTHLAPYTYLFEVSAATTEAGYIDSTIVR 301  
QY 300 TPESVPEGPONCVTGNITGKFSILWDPPTIVTGKFSYRVLYGSPGRILDNSTKOLKF 359  
Db 302 TPESVPEGPONCVTGNITGKFSISWDPPTIVTGKFSYRVLYGSPGRILDNSTKOLRF 361  
QY 360 AFNLTPFTMYDVIYIAAETSAGTGPKSNI SVTPDPVPGAVFDLOLAEVESQVRIWKK 419  
Db 362 AFNLTPFTMYDVIYIAAETSAGTGPKSNI SVTPDPVPGAVFDLOLAEVESQVRIWKK 421  
QY 420 PRQPNGIINQYRVKVLVPTGIIILENTLLTGNNEYINDPMAPEI VNI VEPMVGLYEGSAB 479  
Db 422 PRQPNGIISOYRVKVSIVLSTGVVLENTLLTGDDESISNPMSP EIMNLVDPMIGFYEGSGE 481  
QY 480 MSSDLHSLATFIYNHPDKNFPARNAEDQTSVVTTRNOYITDTAAEOLSVYIRBLVPP 539  
Db 482 MSSDLHSPASFTYNSHPNDPFASTRAEQSSFPVVTTRNOYITDTABEOLSVYVRRLVFP 541  
QY 540 TEHMTSVSFTMGEGPPTVLSVTRQOVPSSIKIINYKNISSSSILLWYDPPPEYNGKI 599  
Db 542 TEHMTSVSFTMGEGPPTVLSVTRQOVPSSIKIINYKNISSSSILLWYDPPPEYNGKI 601  
QY 600 THYTIYABELDNRAFOITIDNSFLITGLKXYTKYKMRVAASHTHGESSLEENDIFVR 659  
Db 602 THYTIYABELDNRAFOITIDNSFLITGLKXYTKYKMRVAASHTHGESSLEENDIFVR 661  
QY 660 TSEDEPESPQDVEVDVTADEIRLKWSPPEKPNGIIIAEVLVYKNIIDLYMKVNTSTDI 719  
Db 662 TSEDEPESPQDVEVDVTADEIRLKWSPPEKPNGIIIAEVLVYKNIIDLYMKVNTSTDI 721  
QY 720 ILRLNRPHTLXNISVRSYTRFGHGNQVSSILLSVRSSETVDPDAPENITVKNISSEGEIS 779  
Db 722 IISDLKPYTLXNISRSYTRLGHGQSSILLSVRSSETVDPDAPENITVKNISSEGEIS 781  
QY 780 FLPPSPNGIISKYTYLXKRSNGNEERTINTTSLTONIKVLKXYTOYIIEVSASTLKGBG 839  
Db 782 FLPPSPNGIISKYTYLXKRSNGNEERTINTTSLTONIKVLKXYTOYIIEVSASTLKGBG 841  
QY 840 VRSAPISILTEADAPSPQDPSVKOLSGVTYKLSWQPLEPENGIIILYTVVWNRSSLK 899  
Db 842 IRSRPSILTEADAPSPQDPSVKOLSGVTYKLSWQPLEPENGIIILYTVVWNRSSLR 901  
QY 900 TTNVTETSELDLDVNVYSVAVTASTRFGDKTGSNIISFQTPGAPSDPPKOVYAN 959  
Db 902 AINATEASVLGDLVDVNDYGACVTASTRFGDNARSSIIINFRTPEGESDPPNDVHYN 961  
QY 960 LSSSSILFWTPPPKENGIIQYYSVYRYNTSGTFMQNFTLHBLTNDPDMNTYSTIIDLKT 1019  
Db 962 LSSSSILFWTPPPKENGIIQYYSVYRYNTSGTFMQNFTLHBLTNDPDMNTYSTIIDLKT 1021  
QY 1020 IFSYTYFWLTASTSVGNKSSDIIEVYTDQDIPGFGVGNLYVESISSTAIVNSVPPAQ 1079  
Db 1022 IFSYTYFWLTASTSVGNKSSDIIEVYTDQDIPGFGVGNLYVESISSTAIVNSVPPSQ 1081  
QY 1080 PNGLVFYVSLIILQOT-PRHVAPPLVYERSYFEDNLEKTYIILKIPSTSEKGSQDVTY 1138  
Db 1082 PNGLVFYVSLIILQOT-PRHVAPPLVYERSYFEDNLEKTYIILKIPSTSEKGSQDVTY 1141  
QY 1139 AOLYIKTEEDVPETSPINTPKNLSTSVLLSWDPVPPKNGAIIISYDLTLQGPENYSFT 1198  
Db 1141 AOLYIKTEEDVPETSPINTPKNLSTSVLLSWDPVPPKNGAIIISYDLTLQGPENYSFT 1199

Db 1142 TOLHIKTEEDVPETSPINTPKNLSTSVLLSWDPVPPKNGAIIISYDLTLQGPENYSFT 1201  
QY 1199 TSDNYIILBELSPFTLYSFFAAARTRKGLGPSSILFFYTDSEVPLAPQNLTLINCTSD 1258  
Db 1202 TSGNHIVILEELSPFTLYSFFAAARTRKGLGPSSILFFYTDSEVPLAPQNLTLINCTSD 1261  
QY 1259 VMLKWSPLPGGIVKVYSFKIHEHETDIYIYKNIISGFKTEAKVGLBEPVSTYSIRVSF 1318  
Db 1262 VMLKWSPLPGGIVKVYSFKIHEHETDIYIYKNIISGFKTEAKVGLBEPVSTYSIRVSF 1321  
QY 1319 TKVGNQFNVVKVETQTESVDDVQNMOCNATSWQSVLVKWDPPKANGIITQWVTV 1378  
Db 1322 TKVGNQFNVVKVETQTESVDDVQNMOCNATSWQSVLVKWDPPKANGIITQWVTV 1381  
QY 1379 RNSTKVSQDQMYTIEKILLANTSYPVKVRASTAGDESBESTCHVSTLPETVPSVPTNIAF 1438  
Db 1382 GNSTKVSQDQMYTIEKILLANTSYPVKVRASTAGDESBESTCHVSTLPETVPSVPTNIAF 1441  
QY 1439 SDVQSTSATLWIRPDTILGYPQNYKITTOQLRAQCKEWESECEVEYQKIQYLYBAHLTE 1498  
Db 1442 SNVQSTSATLWIRPDTILGYPQNYKITTOQLRAQCKEWESECEVEYQKIQYLYBAHLTE 1501  
QY 1499 ETVYGLKFRWYRFOVAASTNAGYGNASNIISTKLPDPPDPPENHVHVSATSPSIS 1558  
Db 1502 ETVYGLKFRWYRFOVAASTNAGYGNASNIISTKLPDPPDPPENHVHVSATSPSIS 1561  
QY 1559 NSEPAVITGPTCYLLIDVKSVDNDEFNISFISKNEENKTIIEIKDLSEIFTRYSVVITAFNG 1618  
Db 1562 NSEPAVITGPTCYLLIDVKSVDNDEFNISFISKNEENKTIIEIKDLSEIFTRYSVVITAFNG 1621  
QY 1619 ISAAVVEGKSSAEMIIVTTLESAPDPPNNMTFQKIPDEVTKFPQLTFLPPSQPNNGIIVVQ 1678  
Db 1622 VSRAVTDGKSSAEMIIVTTLESAPDPPNNMTFQKIPDEVTKFPQLTFLPPSQPNNGIIVVQ 1681  
QY 1679 ALVYREDPTAVQIHNLSIIQKNTFVIAMLEGLKGGTYNISVAVNSAGAKVPKPMRI 1738  
Db 1682 ALVYREDPTAVQIHNLSIIQKNTFVIAMLEGLKGGTYNISVAVNSAGAKVPKPMRI 1741  
QY 1739 TMDIKAPARPKTPPIYDATGKLLAVTSTTIRMPICYSDDHGPIKNVQVLAETGAQ 1798  
Db 1742 TMDIKAPARPKTPPIYDATGKLLAVTSTTIRMPICYSDDHGPIKNVQVLAETGAQ 1801  
QY 1799 HDGNTVYKWDAYFNKARPYFTNEGPNPPCTEGKTKFSGNEBIYIIGADNACMIPGNEK 1858  
Db 1802 HDGNTVYKWDAYFNKARPYFTNEGPNPPCTEGKTKFSGNEBIYIIGADNACMIPGNEK 1861  
QY 1859 ICNGPLKPKQYLFKFRATNMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCLISII 1918  
Db 1862 ICNGPLKPKQYLFKFRATNMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCLISII 1921  
QY 1919 LIGTAFAPARTRQKQEGGTVSPQDAEIIDTKLDOLITVADLEKDERLTR----- 1972  
Db 1922 LIGTAFAPARTRQKQEGGTVSPQDAEIIDTKLDOLITVADLEKDERLTR----- 1981  
QY 1973 ---PISKESFLQHVBEELCTNNLKFQBESELSLPKFLQDLSSTDAOLPMNRAKRPFNK 2029  
Db 1982 SIKPISKESFLQHVBEELCTNNLKFQBESELSLPKFLQDLSSTDAOLPMNRAKRPFNK 2041  
QY 2030 YNNNRVKLIADASVPGSDYINASYISGLCPNEFIATQGPLGTGTFWFMWETRAKT 2089  
Db 2042 Y-NNNRVKLIADASVPGSDYINASYISGLCPNEFIATQGPLGTGTFWFMWETRAKT 2100  
QY 2090 LVMLTQCCEKGRIRCHOYWPENDKPVTGDIIVITKLMEDVOIDMTIRDLKIEREGDCMT 2149  
Db 2101 LVMLTQCCEKGRIRCHOYWPENDKPVTGDIIVITKLMEDVOIDMTIRDLKIEREGDCMT 2160  
QY 2150 VRQCNFTAWPEHGVSPENSAPLIFHFKLVRAHRAHDTTPMIVHCSAGVGRGTGVFIADHLT 2209  
Db 2161 VRQCNFTAWPEHGVSPENSAPLIFHFKLVRAHRAHDTTPMIVHCSAGVGRGTGVFIADHLT 2220  
QY 2210 QHINHDVFDYVGLVAELRSEKRMCMQNLQAYIFLHQICILLDSNKGSNQPCFVNYSL 2269  
Db 2221 QHINHDVFDYVGLVAELRSEKRMCMQNLQAYIFLHQICILLDSNKGSNQPCFVNYSL 2280

QY 2270 QKMSLDMEGDVELEWEETM 2291  
DB 2281 QKMSLDMEGDVELEWEETM 2302

## RESULT 12

US-10-314-232-22  
; Sequence 22, Application US/10314232  
; Publication No. US20030138932A1  
; GENERAL INFORMATION:  
; APPLICANT: MOLLER, NIELS P.H.  
; APPLICANT: MOLLER, KARIN B.  
; APPLICANT: ULLRICH, AXEL  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE  
; FILE REFERENCE: 038602/0686  
; CURRENT APPLICATION NUMBER: US/10/314,232  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/361,096  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: 08/449,609  
; PRIOR FILING DATE: 1995-05-24  
; PRIOR APPLICATION NUMBER: 08/036,210  
; PRIOR FILING DATE: 1995-03-23  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 22  
; LENGTH: 898  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PTPS31-RD#2  
US-10-314-232-22

Query Match 37.8%; Score 4535.5; DB 14; Length 898;  
Best Local Similarity 97.1%; Pred. No. 1.7e-253;  
Matches 861; Conservative 0; Mismatches 1; Indels 25; Gaps 2;  
QY 1429 VPSVPTNIAFSDVQSTSATLTWIRPDTILGYFQNYKITTLQRAQCKEWESECEVYQKI 1488  
DB 13 VPSVPTNIAFSDVQSTSATLTWIRPDTILGYFQNYKITTLQRAQCKEWESECEVYQKI 72  
QY 1489 QYLYEAHLTEETVYGLKFRWYFQVAASNAGYGNASNNWISTKTLPGPPDGPENHVYV 1548  
DB 73 QYLYEAHLTEETVYGLKFRWYFQVAASNAGYGNASNNWISTKTLPGPPDGPENHVYV 132  
QY 1549 ATSPFSISISWSPEAVITGPTCYLIDVKSVDNDEFNIFSKSNEENKTIIEIKDLEIFTRY 1608  
DB 133 ATSPFSISISWSPEAVITGPTCYLIDVKSVDNDEFNIFSKSNEENKTIIEIKDLEIFTRY 192  
QY 1609 SVVITAFGTGNISAAVYEGKSSAEMIVTTLESAPKDPNNMTFOKI PDEVTKFOLTLPSPS 1668  
DB 193 SVVITAFGTGNISAAVYEGKSSAEMIVTTLESAPKDPNNMTFOKI PDEVTKFOLTLPSPS 252  
QY 1669 QPNNGNIQVYQALVYREDDPTAVQIHNLISIIQKNTTFFVIAMLEGLKGHTYNIISVAVNSA 1728  
DB 253 QPNNGNIQVYQALVYREDDPTAVQIHNLISIIQKNTTFFVIAMLEGLKGHTYNIISVAVNSA 312  
QY 1729 GAGPKVPMRITMDIKAPAPKPTPTPIYATGKLLVTSTTITRMPICYSDDHGPKNV 1788  
DB 313 GAGPKVPMRITMDIKAPAPKPTPTPIYATGKLLVTSTTITRMPICYSDDHGPKNV 372  
QY 1789 QVLATETGAQHDGNTVKYDAYFNKARPFTNEGPPNPPCTEGKTFPSGNEEYIIGADN 1848  
DB 373 QVLATETGAQHDGNTVKYDAYFNKARPFTNEGPPNPPCTEGKTFPSGNEEYIIGADN 432  
QY 1849 ACMIPGNEKICNGPLKPKQYLFKFRATNIMQFTDSYSDPKVLGEGLSERTVEIIL 1908  
DB 433 ACMIPGNEKICNGPLKPKQYLFKFRATNIMQFTDSYSDPKVLGEGLSERTVEIIL 492  
QY 1909 SVTLCILSIIILGTAIFAFARINQKQEGTYSPODAEIIDTKLKLDQILITVADLELKD 1968  
DB 493 SVTLCILSIIILGTAIFAFARINQKQEGTYSPODAEIIDTKLKLDQILITVADLELKD 552

QY 1969 RLTR-----PISKKSFLQHVVEELCTNNNLKPFQBEFSELPKF 2004  
DB 553 RLTRYSSPPFRKKEIFVIQLLSYRKSIPISKKSFLQHVVEELCTNNNLKPFQBEFSELPKF 612  
QY 2005 LQDLSTADADLPNRAKNEFPNKPNNNNRVKLIADASVPGSDYINASVISGYLCPNEF 2064  
DB 613 LQDLSTADADLPNRAKNEFPNKPNNNNRVKLIADASVPGSDYINASVISGYLCPNEF 671  
QY 2065 IATQGPLCTGVDGFWRMVWETRAKTLVMLTQCFEKGIRIRCHOYWPEDNKPVTVFGDIVIT 2124  
DB 672 IATQGPLCTGVDGFWRMVWETRAKTLVMLTQCFEKGIRIRCHOYWPEDNKPVTVFGDIVIT 731  
QY 2125 KLMEDVQIDWTIRDLKIERHGDCTVTRQCNFTAMPBHGVPENSAPLIHFVKLVASRAHD 2184  
DB 732 KLMEDVQIDWTIRDLKIERHGDCTVTRQCNFTAMPBHGVPENSAPLIHFVKLVASRAHD 791  
QY 2185 TTPMIVHCSAGVGRGTGFIALDHLTOHINDHDPVDIYGLVAELSRMCMQVONLAQYIFL 2244  
DB 792 TTPMIVHCSAGVGRGTGFIALDHLTOHINDHDPVDIYGLVAELSRMCMQVONLAQYIFL 851  
QY 2245 HQCILDLNLSKGSNOPICFVNYSAQKMSLDAMEGDVELEWEETM 2291  
DB 852 HQCILDLNLSKGSNOPICFVNYSAQKMSLDAMEGDVELEWEETM 898

## RESULT 13

US-10-087-684-8  
; Sequence 8, Application US/10087684  
; Publication No. US2004002916A1  
; GENERAL INFORMATION:  
; APPLICANT: Edinger, Shlomit R.  
; APPLICANT: MacDougall, John R.  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Stone, David J.  
; APPLICANT: Grosse, William M.  
; APPLICANT: Lepley, Denise M.  
; APPLICANT: Rieger, Daniel K.  
; APPLICANT: Burgess, Catharine E.  
; APPLICANT: Casman, Stacie, J.  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Boldog, Ferenc L.  
; APPLICANT: Li, Li  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Mishra, Vishnu  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Zerhusen, Bryan D.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Miller, Charles E.  
; APPLICANT: Gangolli, Esha A.  
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
; FILE REFERENCE: 21402-214 CIP  
; CURRENT APPLICATION NUMBER: US/10/087,684  
; CURRENT FILING DATE: 2003-03-10  
; PRIOR APPLICATION NUMBER: 60/253,834  
; PRIOR FILING DATE: 2000-11-29  
; PRIOR APPLICATION NUMBER: 60/250,926  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: 60/264,180  
; PRIOR FILING DATE: 2001-01-25  
; PRIOR APPLICATION NUMBER: 60/274,194  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/313,656  
; PRIOR FILING DATE: 2001-08-20  
; PRIOR APPLICATION NUMBER: 60/327,456  
; PRIOR FILING DATE: 2001-10-05  
; NUMBER OF SEQ ID NOS: 220  
; SOFTWARE: CuraSeqList version 0.1

```
; SEQ ID NO 8
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-087-684-8

Query Match      37.4%; Score 4488; DB 15; Length 855;
Best Local Similarity 99.6%; Pred. No. 9e-251;
Matches 850; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1053 PEGFVGNLTYESISSTAINVSWVPPAQNGLVYVYVLSLILOQTPRHVRPPLVYERSIYF 1112
Db      |||||
QY      3 PEGFVGNLTYESISSTAINVSWVPPAQNGLVYVYVLSLILOQTPRHVRPPLVYERSIYF 62
Db      |||||
QY 1113 DNLEKTYDILKTPSTEGFSDTYTAQLYIKTEEDVPETSPINTFKNLSSTVLLSWD 1172
Db      |||||
QY 63 DNLEKTYDILKTPSTEGFSDTYTAQLYIKTEEDVPETSPINTFKNLSSTVLLSWD 122
Db      |||||
QY 1173 PPVKPENGAIISYDLTLQGPENYSFITSNDNYIILELSPFTLYSPFAAARTKGLGPSSI 1232
Db      |||||
QY 123 PPVKPENGAIISYDLTLQGPENYSFITSNDNYIILELSPFTLYSPFAAARTKGLGPSSI 182
Db      |||||
QY 1233 LFFYTDESVPPLAPPQNLTINCTSDFVWLKWSPLPGGIVKVSFKIHEHETDTIYKN 1292
Db      |||||
QY 183 LFFYTDESVPPLAPPQNLTINCTSDFVWLKWSPLPGGIVKVSFKIHEHETDTIYKN 242
Db      |||||
QY 1293 ISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNGNQFSNVVKFTTQESVPDVQNMCMATS 1352
Db      |||||
QY 243 ISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNGNQFSNVVKFTTQESVPDVQNMCMATS 302
Db      |||||
QY 1353 WSVLVKWDPPPKANGIIQYVMTVERNSTKVSPODHMTFIKLLANTSYVFKVRASTSA 1412
Db      |||||
QY 303 WSVLVKWDPPPKANGIIQYVMTVERNSTKVSPODHMTFIKLLANTSYVFKVRASTSA 362
Db      |||||
QY 1413 GGEDESTCHVSTLPETVPSPVNTIAFSDVQSTSATLTWIRPDTILGYFQNYKITTLQRAQ 1472
Db      |||||
QY 363 GGEDESTCHVSTLPETVPSPVNTIAFSDVQSTSATLTWIRPDTILGYFQNYKITTLQRAQ 422
Db      |||||
QY 1473 KCKEWESEBCEVYQKIYLYEAHLTEETVYGLKFRWYFQVAASTNAGYGNASNIWSTK 1532
Db      |||||
QY 423 KCKEWESEBCEVYQKIYLYEAHLTEETVYGLKFRWYFQVAASTNAGYGNASNIWSTK 482
Db      |||||
QY 1533 TLPGPPDGPENVHVATGPFSSISWSBPAVITGTCYLIDVKSVDNDEFNISFTKSNE 1592
Db      |||||
QY 483 TLPGPPDGPENVHVATGPFSSISWSBPAVITGTCYLIDVKSVDNDEFNISFTKSNE 542
Db      |||||
QY 1593 ENKTIKLEIFTRYSVVITAFPTGNISAAAYVEGKSSAEMIVTTLESAPKPPNNMTFOK 1652
Db      |||||
QY 543 ENKTIKLEIFTRYSVVITAFPTGNISAAAYVEGKSSAEMIVTTLESAPKPPNNMTFOK 602
Db      |||||
QY 1653 IPDEVTKFQITLPPSQPNQNTQVYQALVYREDDPTAVQIHNLISIIQKNTFTVIAMLEGL 1712
Db      |||||
QY 603 IPDEVTKFQITLPPSQPNQNTQVYQALVYREDDPTAVQIHNLISIIQKNTFTVIAMLEGL 662
Db      |||||
QY 1713 KSGHTYNI SVYAVNSAGAKVPKMRITMDIKAPAREKTKPTPIYDATGKLLVTSITITIR 1772
Db      |||||
QY 663 KSGHTYNI SVYAVNSAGAKVPKMRITMDIKAPAREKTKPTPIYDATGKLLVTSITITIR 722
Db      |||||
QY 1773 MPICYYSDDHGPIKNVQVLATETGAQHDGNVTKWYDAYFNKARPYFTNFGFNPCCTEGK 1832
Db      |||||
QY 723 MPICYYSDDHGPIKNVQVLATETGAQHDGNVTKWYDAYFNKARPYFTNFGFNPCCTEGK 782
Db      |||||
QY 1833 TKFSGNEEYIIGADNACMI PGNEDKI CNGPLKPKQYLFKFRATNIMGQFTDSYSDPV 1892
Db      |||||
QY 783 TKFSGNEEYIIGADNACMI PGNEDKI CNGPLKPKQYLFKFRATNIMGQFTDSYSDPV 842
Db      |||||
QY 1893 KTLGEGLSERTVE 1905
Db      |||||
QY 843 KTLGEGLSERTLE 855
Db      |||||

RESULT 14
US-10-218-779-8
```

```
; Sequence 8, Application US/10218779
; Publication NO. US20040029222A1
; GENERAL INFORMATION:
; APPLICANT: Edinger, Shlomit
; APPLICANT: Macbougall, John
; APPLICANT: Millet, Isabelle
; APPLICANT: Ellerman, Karen
; APPLICANT: Stone, David
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William
; APPLICANT: Alsobrook II, John
; APPLICANT: Lepley, Denise
; APPLICANT: Rieger, Daniel
; APPLICANT: Burgess, Catherine
; APPLICANT: Casman, Stacie
; APPLICANT: Spytek, Kimberly
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Mishra, Vishnu
; APPLICANT: Patturajan, Meera
; APPLICANT: Shenoy, Suresh
; APPLICANT: Rastelli, Luca
; APPLICANT: Tchernev, Velizar
; APPLICANT: Vernet, Corine
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Malyankar, Uriel
; APPLICANT: Guo, Xiaojia
; APPLICANT: Miller, Charles
; APPLICANT: Gangolli, Bsha
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-214
; CURRENT APPLICATION NUMBER: US/10/218,779
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 60/253,834
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/250,-926
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/264,180
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/313,656
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/327,456
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-218-779-8

Query Match      37.4%; Score 4488; DB 15; Length 855;
Best Local Similarity 99.6%; Pred. No. 9e-251;
Matches 850; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1053 PEGFVGNLTYESISSTAINVSWVPPAQNGLVYVYVLSLILOQTPRHVRPPLVYERSIYF 1112
Db      |||||
QY      3 PEGFVGNLTYESISSTAINVSWVPPAQNGLVYVYVLSLILOQTPRHVRPPLVYERSIYF 62
Db      |||||
QY 1113 DNLEKTYDILKTPSTEGFSDTYTAQLYIKTEEDVPETSPINTFKNLSSTVLLSWD 1172
Db      |||||
QY 63 DNLEKTYDILKTPSTEGFSDTYTAQLYIKTEEDVPETSPINTFKNLSSTVLLSWD 122
Db      |||||
QY 1173 PPVKPENGAIISYDLTLQGPENYSFITSNDNYIILELSPFTLYSPFAAARTKGLGPSSI 1232
Db      |||||
QY 123 PPVKPENGAIISYDLTLQGPENYSFITSNDNYIILELSPFTLYSPFAAARTKGLGPSSI 182
Db      |||||
QY 1233 LFFYTDESVPPLAPPQNLTINCTSDFVWLKWSPLPGGIVKVSFKIHEHETDTIYKN 1292
Db      |||||
QY 183 LFFYTDESVPPLAPPQNLTINCTSDFVWLKWSPLPGGIVKVSFKIHEHETDTIYKN 242
Db      |||||
QY 1293 ISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNGNQFSNVVKFTTQESVPDVQNMCMATS 1352
Db      |||||
```

Db 243 ISGFKTEAKLVGLPEPSTYSIRSAFTKVGNGQFSSNVKFTTQESVDPVQNMCMATS 302  
Qy 1353 WQSVLVKWDPPKANGIIITQYMTVVERNSKVSQDQHMVTFIKLLANTSIVFKVRASTSA 1412  
Db 303 WQSVLVKWDPPKANGIIITQYMTVVERNSKVSQDQHMVTFIKLLANTSIVFKVRASTSA 362  
Qy 1413 GEGDESTCHVSTLPETVPSVPTNIAPSDVOSTSATLTWRPDTILGYFQNYKITTLQRAQ 1472  
Db 363 GEGDESTCHVSTLPETVPSVPTNIAPSDVOSTSATLTWRPDTILGYFQNYKITTLQRAQ 422  
Qy 1473 KCKEWESEECVYQKIQLYEAHLTETVYGLKFRWYFQVAASNAGVGNASNIWK 1532  
Db 423 KCKEWESEECVYQKIQLYEAHLTETVYGLKFRWYFQVAASNAGVGNASNIWK 482  
Qy 1533 TLPGPPDGPENVHVATSPFISISWSBPAVITGPTCYLIDVKSVDNDEFNIFSIKSNE 1592  
Db 483 TLPGPPDGPENVHVATSPFISISWSBPAVITGPTCYLIDVKSVDNDEFNIFSIKSNE 542  
Qy 1593 ENKTIKOLEIFTRYSVVTATGNISAAVYVSGKSAEMIVTTLSSAPKDPNNMTFQK 1652  
Db 543 ENKTIKOLEIFTRYSVVTATGNISAAVYVSGKSAEMIVTTLSSAPKDPNNMTFQK 602  
Qy 1653 IPDEVTKFOLTFLPBPSPNGNIQVQALVYREDPTAVQIHNLSIIQKNTFVIAMLEGL 1712  
Db 603 IPDEVTKFOLTFLPBPSPNGNIQVQALVYREDPTAVQIHNLSIIQKNTFVIAMLEGL 662  
Qy 1713 KGGHTYNI SYAVNSAGAGKVPWRITMDIKAPARKTPTPIYDATGKLLVTSTTITIR 1772  
Db 663 KGGHTYNI SYAVNSAGAGKVPWRITMDIKAPARKTPTPIYDATGKLLVTSTTITIR 722  
Qy 1773 MPICYSDHGHPIKNVQVLATETGAQHDGNTWKYDAYFNKARPYTNEGFPNPPCTEGK 1832  
Db 723 MPICYSDHGHPIKNVQVLATETGAQHDGNTWKYDAYFNKARPYTNEGFPNPPCTEGK 782  
Qy 1833 TKESGNEEYIIIGADNACMIPGNEKICNGPLKPKQYLFKFRATNIMGQFTDSDYSDPV 1892  
Db 783 TKESGNEEYIIIGADNACMIPGNEKICNGPLKPKQYLFKFRATNIMGQFTDSDYSDPV 842  
Qy 1893 KTLGEGLSERTVE 1905  
Db 843 KTLGEGLSERTLE 855

RESULT 15  
US-10-314-232-15  
; Sequence 15, Application US/10314232  
; Publication No. US20030138932A1  
; GENERAL INFORMATION:  
; APPLICANT: MOLLER, NIELS P.H.  
; APPLICANT: MOLLER, KARIN B.  
; APPLICANT: ULLRICH, AXEL  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE  
; FILE REFERENCE: 038602/0686  
; CURRENT APPLICATION NUMBER: US/10/314,232  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/361,096  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: 08/449,609  
; PRIOR FILING DATE: 1995-05-24  
; PRIOR APPLICATION NUMBER: 08/036,210  
; PRIOR FILING DATE: 1995-03-23  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 401  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-314-232-15

Query Match 15.9%; Score 1902.5; DB 14; Length 401;  
Best Local Similarity 92.8%; Pred. No. 1.1e-101;  
Matches 361; Conservative 2; Mismatches 1; Indels 25; Gaps 2;

Qy 1927 FARIROKQKEGGTYSPQDAEIIIDTKLQDLITVADLELKDRLTR----- 1972  
Db 14 YSEIROKQKEGGTYSPQDAEIIIDTKLQDLITVADLELKDRLTRYSSFFFRKEIFVI 73  
Qy 1973 -----PISKKSFLQHVVELCTNNLKFQEESELPKFLQDLSSTADLPMNRKN 2022  
Db 74 QLLSVKSKIKPISKKSFLQHVVELCTNNLKFQEESELPKFLQDLSSTADLPMNRKN 133  
Qy 2023 RFPNIKPYNNNRVVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTVDGFWRMV 2082  
Db 134 RFPNIKPYNNNRVVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTVDGFWRMV 192  
Qy 2083 WETRAKTLVMTQCFEKGIRCHQYWPEDNKPVTVPDGIIVITKLMEDVQIDMTIRDLKIE 2142  
Db 193 WETRAKTLVMTQCFEKGIRCHQYWPEDNKPVTVPDGIIVITKLMEDVQIDMTIRDLKIE 252  
Qy 2143 RHGDCMTVQCNFTAWPEHGVSPENSAPLTHFVKLVRASRAHDTTPMIVHCSAGVGRGVF 2202  
Db 253 RHGDCMTVQCNFTAWPEHGVSPENSAPLTHFVKLVRASRAHDTTPMIVHCSAGVGRGVF 312  
Qy 2203 IALDHLTQHINDHDFVDIYGLVAELSERMCMVQNLAQYIFLHQICILDLILSNKGSNPIC 2262  
Db 313 IALDHLTQHINDHDFVDIYGLVAELSERMCMVQNLAQYIFLHQICILDLILSNKGSNPIC 372  
Qy 2263 FVNSALQKQDSLDAMEGDVELEWETTMM 2291  
Db 373 FVNSALQKQDSLDAMEGDVELEWETTMM 401

Search completed: June 30, 2005, 15:20:57  
Job time : 251 secs

**THIS PAGE BLANK (USPTO)**